

Semojum: A Multimodal AI-Based Braille Transcription Support Service for Visual Materials and Layouts

Taemin Kim*, Junhyeok Lee**, Haeun Cho***, Hyunju Kim****, Sanghyun Kim*****, Jimin Lee*****,
Won Joo Lee†

*Student, School of Computing, KAIST, Daejeon, Korea

**Student, Dept. of Computer Engineering, Konkuk University, Seoul, Korea

***Student, Dept. of Computer Science, Dongduk Women's University, Seoul, Korea

****Student, School of Robotics, Kwangwoon University, Seoul, Korea

*****Student, School of Electronic Engineering, Hanyang University, Seoul, Korea

*****Student, Dept. of AI, Big Data & Management from Kookmin University, Seoul, Korea

†Professor, Dept. of Computer Science & Engineering, Inha Technical College, Incheon, Korea

[Abstract]

In this paper, we propose a Semojum, a multimodal AI-based braille translation support service designed to provide real-time braille books for the visually impaired. Semo-Jeom consists of a DeepSeek-OCR-2 based layout analysis module, a Semantic NMS algorithm, a table layout optimization engine, and a GPT-4o based image captioning module. The layout analysis module scales page images up to 1,536 pixels and outputs element types, OCR text, bounding boxes, and reading orders in a single pass using specialized markup tokens. To refine the output, the Semantic NMS algorithm determines redundancy based on the semantic inclusion relationships between extracted text contents. The table layout optimization engine extracts structured data using the JSON Schema strict output of GPT-4o-mini Vision. For visual accessibility, the image captioning module utilizes GPT-4o Vision, injecting adjacent text from the preceding and following pages as context; it then classifies images into seven distinct patterns and applies specific descriptive strategies for each. Furthermore, Semo-Jeom implements a hybrid braille translation engine that combines the Brailify library with a custom LaTeX tokenizer to support the 2024 Revised Korean Mathematics Braille Regulations. By integrating a Human-in-the-loop structure that presents AI outputs as verifiable drafts and a tab-based three-mode interface, the system significantly enhances the operational efficiency of professional braille translators.

▶ **Key words:** Braille Translation Automation, Multimodal AI, Layout Analysis, Korean Mathematical Braille, Human-in-the-loop, BRFI

• First Author: Taemin Kim, Corresponding Author: Won Joo Lee

*Taemin Kim (arminus0402@gmail.com), School of Computing, KAIST

**Junhyeok Lee (rugs12345@gmail.com), Dept. of Computer Engineering, Konkuk University

***Haeun Cho (johaeun0802@naver.com), Dept. of Computer Science, Dongduk Women's University

****Hyunju Kim (hjyg010s@gmail.com), School of Robotics, Kwangwoon University

*****Sanghyun Kim (mordecia3800@gmail.com), School of Electronic Engineering, Hanyang University

*****Jimin Lee (dlwals922@naver.com), Dept. of AI, Big Data & Management from Kookmin University

†Won Joo Lee (wonjoo2@inhac.ac.kr), Dept. of Computer Science & Engineering, Inha Technical College

• Received: 2026. 05. 20, Revised: 2026. 06. 15, Accepted: 2026. 06. 15.

[요 약]

본 논문에서는 시각장애인들에게 실시간 점역 도서 공급을 위한 멀티모달 AI 기반 점역 지원 서비스인 세모점을 제안한다. 이 세모점은 DeepSeek-OCR-2 기반 레이아웃 분석, Semantic NMS 알고리즘, 표 레이아웃 최적화 엔진, GPT-4o 기반 이미지 캡셔닝 모듈로 구성된다. DeepSeek-OCR-2 기반 레이아웃 분석 모듈은 페이지 이미지를 최대 1,536픽셀로 스케일링한 뒤 전용 마크업 토큰으로 요소 유형, OCR 텍스트, 바운딩 박스, 독서 순서를 단일 패스로 출력한다. Semantic NMS 알고리즘은 OCR로 추출된 텍스트 내용 간 포함 관계를 기준으로 중복 여부를 판단한다. 표 레이아웃 최적화 엔진은 먼저 GPT-4.1 mini Vision의 JSON Schema strict 출력으로 표 구조를 추출한다. GPT-4o 기반 이미지 캡셔닝 모듈은 GPT-4o Vision을 사용하며, 이미지 앞뒤 인접 텍스트를 문맥으로 주입하고 7가지 패턴으로 분류한 뒤 패턴별 서술 전략을 적용한다. 또한 세모점은 Braillify 라이브러리와 커스텀 LaTeX 토크나이저를 결합하여 2024년 개정 한국 수학 점자 규정을 지원하는 하이브리드 점역 엔진을 구현한다. AI 출력을 검수 가능한 초안으로 제시하는 Human-in-the-loop 구조와 탭 기반 3모드 인터페이스를 통해 점역사의 작업 효율을 지원한다.

▶ **주제어:** 점역 자동화, 멀티모달 AI, 레이아웃 분석, 한국 수학 점자, Human-in-the-loop, BRFI

Introduction

디지털 출판 시대에 시각장애인은 점자 도서 발행이 어렵기 때문에 신간을 실시간으로 접하기 어렵다. 국립장애인도서관 통계에 따르면 일반 도서 한 권이 점자 도서로 제작되기까지 평균 5.3개월이 소요되며, 연간 생산량은 약 600권으로 전체 출판 도서의 1% 미만에 불과하다[1]. 특히 수식·도표·그래프가 혼재된 이공계 교재는 한 페이지 점역에 수 시간이 소요되어, 시각장애 대학생의 학습 자료 접근은 만성적으로 지연되고 있다[2]. 이러한 교재 공급 부족의 핵심은 고숙련 점역 인력이 절대적으로 부족하다는 것이다. 국내 점역사는 약 1,500명이나, 수학·과학 등 다양한 과목 점역이 가능한 인력은 약 400명 정도이다[3]. 기존 디지털 도구들도 이 문제를 해결하지 못한다. SENSEE는 텍스트 점역에 특화되어 있어 수식 변환과 표 처리가 불가하고, dot 패드는 촉각 디스플레이 전용 하드웨어로 수식을 일반 텍스트와 동일하게 처리하여 이공계 자료에 적합하지 않다[4, 5]. 이러한 공통적인 한계는 레이아웃 분석과 수식·표·이미지 처리를 단일 파이프라인으로 통합하는 도구가 없기 때문이다.

이러한 한계를 최소화하기 위해 본 논문에서는 멀티모달 AI 기반의 점자 전자 지원 서비스인 세모점을 제안한다. 이 세모점은 DeepSeek-OCR-2 기반 레이아웃 분석, Semantic NMS, 규칙 기반 표 레이아웃 최적화, GPT-4o 기반 이미지 캡셔닝을 단일 파이프라인으로 통합하여 이공계 교재의 복합 문서 요소를 일관된 프로세스로 처리한

다. 또한 Braillify와 커스텀 LaTeX 파서를 결합한 하이브리드 점역 엔진으로 2024년 개정 한국 수학 점자 규정을 지원하며, AI 결과를 초안으로 제시하고 점역사가 매 단계에서 검수·수정하는 Human-in-the-loop 구조로 설계한다. 본 논문 2장에서 기존 서비스를 벤치마킹하고, 관련 기술을 검토한다. 3장에서 시스템을 설계하고, 4장에서 핵심 모듈 구현 방법 및 구현 결과에 대하여 설명한다. 그리고 5장에서 결론을 맺는다.

II. Preliminaries

1. Bench-marking

기존 점역 관련 서비스를 기능 단위로 분석하고, 본 논문에서 제안하는 세모점과 비교한 결과는 Table 1과 같다.

Table 1. Functional comparison of Braille translation services

Comparison Item	Manual	SENSEE	dot
Text	○	○	○
Equations (LaTeX)	○	×	△
Images	○	△	△
Tables	○	×	×
Multi-document	○	×	×
Braille special symbols	○	×	×
Automated Processing	×	△	△

Table 1에서 수작업 점역은 수십 년간 검증된 현행 기준 방법(Baseline)으로 정확도는 높으나, 권당 수개월의 작업 기간과 고숙련 인력이 필요하다. SENSEE는 텍스트 점역에 특화되어 있어 수식·표·다단 문서 처리와 점역 기호 삽입을 지원하지 않는다[4]. dot 패드는 촉각 디스플레이 전용 하드웨어로 수식을 구분 없이 처리하고 소프트웨어 점역 기능은 제한적이다[5].

2. Related works

1) Document layout analysis

DocLayout-YOLO[6]는 YOLO 계열 탐지기를 문서에 특화하여 높은 바운딩 박스 정확도를 제공하지만, 별도의 OCR 파이프라인과 독서 순서 결정 후처리가 필요하다. LayoutLMv3[7]은 텍스트-비전 통합 문서 이해 모델로 영어 문서에서 높은 성능을 보이나, 한국어 수식·도표 혼재 문서에서의 일반화가 제한적이다. DeepSeek-OCR-2[8]는 레이아웃 감지·OCR·독서순서 결정을 단일 추론으로 통합하여, 후처리 파이프라인 없이 다단 문서의 시맨틱 독서 순서를 직접 제공한다.

2) Math braille conversion

영어권에서는 Nemeth Code 및 UEB 규격을 지원하는 Liblouis[9]가 표준 도구로 사용되지만, 2024년 전면 개정된 한국 수학 점자 규정[10]에 대한 지원은 현재 어느 도구에서도 제공되지 않는다. 특히 분수·적분·극한 등 다계층 수식에서 요구되는 재귀적 그룹화 표기는 규칙 기반 파서 없이는 처리할 수 없어 커스텀 구현이 불가피하다.

3) Large multimodal language model

GPT-4o[11]를 비롯한 대형 비전-언어 모델은 Few-Shot 프롬프트로 도메인에 최적화된 출력을 유도할 수 있다. OpenAI Responses API의 JSON Schema strict 모드는 응답 포맷을 강제하여, 다단 헤더와 병합 셀이 포함된 표에서 파싱 오류를 원천 차단한다[11]. 이 기능은 표 구조 추출의 신뢰성을 확보하는 데 직접 활용되었다.

레이아웃 감지부터 BRF 초안 생성까지 전 과정을 자동 처리하고, 결과물을 완성본이 아닌 검수 가능한 초안으로 제시한다. 처리 대상 요소는 텍스트, 수식(LaTeX), 표, 이미지의 네 가지로, 각 유형은 특화된 파이프라인을 통해 처리된다. 세모점은 3 가지 처리 모드를 탭 기반 인터페이스로 제공한다. 첫째, 초안 생성 모드는 PDF·HWP 등 복합 문서를 입력받아 OCR·레이아웃 분석·이미지 캡처·표 최적화를 순차 실행하고, 점역 직전의 검수 가능한 텍스트 초안을 실시간으로 스트리밍한다. 둘째, 텍스트 점역 모드는 이미 정제된 텍스트를 입력받아 OCR·레이아웃 단계 없이 점역 엔진에 직접 전달하여 빠른 응답 속도를 보장한다. 셋째, 이미지 점역 모드는 두 모드를 단일 파이프라인으로 통합한 end-to-end 모드로, 이미지 입력부터 BRF 파일 출력까지 전 과정을 자동 처리한다.

2. System Architecture

전체 시스템 구조는 Fig. 1과 같다.

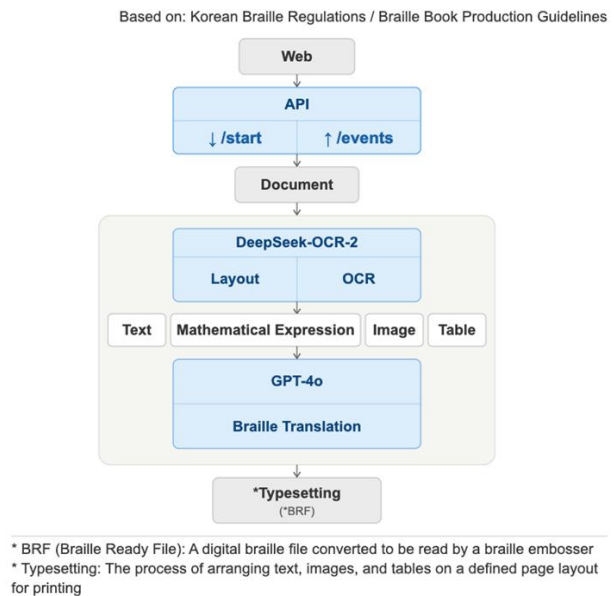


Fig. 1. System Architecture

Fig. 1의 세모점은 React 웹 클라이언트와 FastAPI AI 서버의 2-tier 구조이다. 파일 업로드와 작업 생성에는 REST 방식으로, 페이지별 처리 결과의 실시간 전달에는 SSE를 적용한다. 그리고 점역사가 전체 처리 완료를 기다리지 않고 완료된 페이지부터 실시간으로 확인할 수 있다. GPU 추론 단계는 asyncio.Lock으로 순차 실행하여 동시 요청에서도 자원 경쟁 없이 처리된다.

III. Design of Semojum

1. Design Principles and Processing Modes

본 논문에서 제안하는 세모점은 멀티모달 AI 기반의 점자 전사 지원 서비스이다. 세모점 설계의 핵심 원칙은 AI 자동화와 전문가 검수의 역할 분리이다. AI 파이프라인은

3. Backend API and AI Pipeline

세모점 서버는 FastAPI 기반의 단일 서버로 구성하고, HTTPS 및 SSE 통신을 지원한다. 서버는 이미지에서 점역 직전 텍스트를 추출하는 초안 생성 모드, 텍스트 파일을 점자로 변환하는 점자 번역(텍스트) 모드, 이미지 입력부터 점자 출력까지 전 단계를 통합 처리하는 점자 번역(이미지) 모드의 세 가지 변환 모드를 지원한다. 클라이언트가 업로드 시 지정한 모드에 따라 파이프라인을 분기하여 실행한다. 서버-클라이언트 통신은 두 가지 엔드포인트로 구성한다. '/start' 엔드포인트는 파일 업로드 및 작업 시작을 처리하며, 클라이언트 IP를 기반으로 고유한 job_id를 즉시 발급하여 백그라운드에서 파이프라인을 실행한다. '/events' 엔드포인트는 SSE 스트리밍을 통해 페이지 단위 변환 결과를 실시간으로 전송하여 점역사 전체 변환이 완료되기 전에 처리된 페이지부터 순차적으로 확인 및 편집할 수 있다.

AI 파이프라인은 문서 요소 유형에 따라 텍스트, 수식, 이미지, 표로 분류하여 각각 처리한다. AI 파이프라인 구성은 Table 2와 같다.

Table 2. AI pipeline configuration of Semojum

Processing Stage	Model / Tool
Layout Detection·OCR·Reading Order	DeepSeek-OCR-2
Text OCR Correction	GPT-4o
Image captioning	GPT-4o Vision
Table extraction	GPT-4.1 mini Vision
Table layout	Rule-based engine
Text braille	Braillify + Tokenizer
BRF formatting	Rule-based formatter

Table 2에서는 DeepSeek-OCR-2를 통해 레이아웃 분석, 읽기 순서 부여, OCR을 통합적으로 수행하며, 환각 방지를 위한 중복 방지 로직과 읽기 순서 보존을 적용한다. 이미지는 GPT-4o Vision을 활용하여 시각장애인 음성 데이터셋 기반 프롬프트 엔지니어링과 앞뒤 문맥 주입을 통해 설명문을 생성하고, 표는 점역사 페르소나 기반 프롬프트 엔지니어링으로 복잡한 표 구조를 보존하며 셀 내용의 미 분석을 통해 레이아웃 모드를 자동 판별한다. 이후 변환된 텍스트는 2024 한국 점자 규정이 적용된 한글 점역 엔진과 정규표현식 기반 특수 기호 점자 매핑을 통해 점자로 변환되며, 2025 점자 도서 제작 지침에 맞는 들여쓰기 및 태깅 처리가 적용된 BRF 형식으로 출력된다.

IV. Implementation of Semojum

1. Core Module of Semojum

1) Layout Detection and Semantic NMS

DeepSeek-OCR-2는 페이지 이미지를 최대 1,536픽셀로 스케일링한 뒤 전용 마크업 토큰으로 요소 유형, OCR 텍스트, 바운딩 박스, 독서 순서를 단일 패스로 출력한다. 다단-혼합 레이아웃 문서에서는 동일 요소가 중복 검출되는 경우가 발생하며, 기하학적 좌표만으로 재정렬하면 모델이 파악한 문맥적 독서 흐름이 왜곡된다. 이를 방지하기 위해 Semantic NMS를 적용한다. 이는 OCR로 추출된 텍스트 내용 간 포함 관계를 기준으로 중복 여부를 판단한다. Semantic NMS 알고리즘은 Fig. 2와 같다.

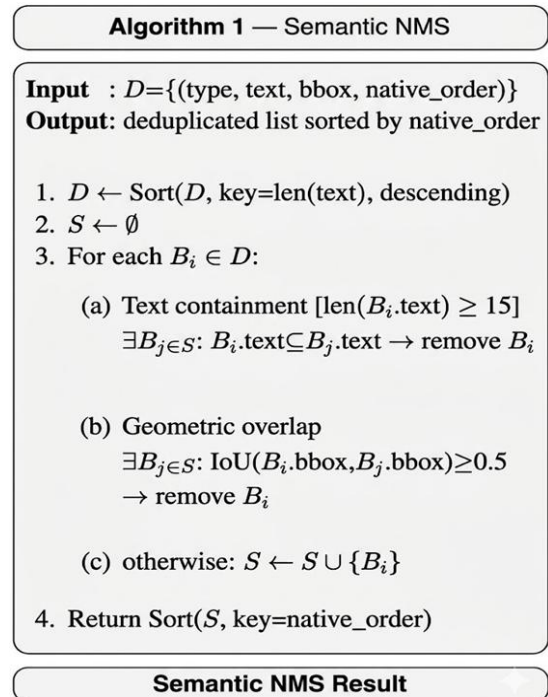


Fig. 2. Semantic NMS Algorithm

2) Mathematical Braille Conversion Engine

기존 점역 라이브러리-영어 수학 점자의 표준 도구인 Liblouis[9]와 한국어 점역 라이브러리 Braillify는 2024년 개정 한국 수학 점자 규정에 따른 LaTeX 수식 변환 기능을 제공하지 않는다[12]. 이 공백을 메우기 위해 하이브리드 구조의 커스텀 엔진을 설계한다.

엔진의 처리 과정은 먼저 정규식으로 인라인 수식 구간과 일반 텍스트를 분리한다. 일반 텍스트는 Braillify 라이브러리로 한국어 점역하고, 영문자·숫자는 별도 매핑 테이블로 보완한다. 수식 구간은 커스텀 LaTeX 토큰라이저가 30종 이상의 토큰을 우선순위 순서로 매칭하여 재귀적으

로 점자로 변환한다. 그룹화 필요 여부는 `needs_grouping()` 함수가 판단한다. 단일 자/숫자 또는 그리스 기호는 그룹화를 생각하고, 2자 이상 표현이거나 연산자를 포함하면 점자 괄호(⠠⠠⠠⠠⠠⠠)를 자동 삽입한다. 분수는 한국어 읽기 순서인 'b분의 a'에 따라 분모 점역값을 먼저 출력한 뒤 분수선(⠨)과 분자 점역값을 배치한다.

3) Table Layout Optimization

표 레이아웃 최적화 엔진은 먼저 GPT-4.1 mini Vision의 JSON Schema strict 출력으로 표 구조(헤더 행·데이터 행·계층 깊이)를 추출한다. 이후 규칙 기반 `_decide_layout()` 함수가 Table 4의 우선순위 순서에 따라 출력 모드를 결정한다. 폭 추정식은 $\max_w = \Sigma(\text{열별 최대 점자 폭}) + 2 \times (\text{열 수} - 1)$ 이다. Transposed 모드 적용 시 "표의 이해를 돕기 위해 행과 열을 바꾸어 적었습니다.", Numbering 모드 적용 시 "표를 번호 체계로 풀어 적었습니다."라는 점역자 주가 BRF에 자동 삽입된다.

4) LLM Optimization Strategy

텍스트 OCR 교정에는 GPT-4o의 temperature를 0으로 적용한다. 이는 동일 입력에 대해 항상 동일한 출력을 생성하여 결과의 결정론적 재현성을 확보한다. 핵심 설계 과제는 LLM의 맥락 기반 대체 경향을 억제하는 것이다. Fig. 3과 같이 "시각적으로 유사한 문자 간 오인식만 교정하고 내용을 변경하지 말 것"을 시스템 프롬프트에 명시하고 실제 OCR 오류 패턴을 Few-Shot 예시로 포함한다.

[System] Post-processing AI for OCR.
Only correct visually misrecognized errors
and do not arbitrarily alter the content.

[Few-Shot]

입력: "간간한 LLM"
→ 출력: "간간한 LLM"
→ 입력: "cは c는 광속이다"

temperature: 0.0 | model: gpt-4o

Fig. 3. OCR correction prompt (Few-Shot design)

5) Image Captioning Strategy

이미지 캡셔닝은 GPT-4o Vision을 사용하며, 이미지 앞뒤 인접 텍스트를 문맥으로 주입하고 7가지 패턴으로 분류한 뒤 패턴별 서술 전략을 적용한다. DIAGRAM 패턴에는 시각적 형태 묘사 대신 해당 분야의 과학적·수학적 원리를 직접 서술하도록 지시하여 교육적 맥락에 부합하는 점역자 주(시각자료 대체 텍스트)를 생성한다. 7가지 패턴 분류 프롬프트 구조는 Fig. 4와 같다.

[System] Braille transcription AI.
Classify the image into one of 7 patterns:
COMIC | FLOWCHART | DIAGRAM | SPATIAL | TIMELINE | MULTI_PHOTO | GENERAL

[Context] Adjacent text (CONTEXT_RANGE=1):
before: {context_before}
after : {context_after}

[Strategy per pattern]
DIAGRAM → state scientific/mathematical principles (not visual form)
FLOWCHART → describe process flow and decision branches
SPATIAL → describe relative positions and spatial layout
TIMELINE → describe chronological sequence and milestones
(others) → describe content appropriate to educational context

temperature: 0.0 | model: gpt-4o-vision

Fig. 4. Image captioning 7-pattern classification prompt

6) BRF Formatting

BRF Formatter는 32칸×25줄 제약을 적용하며, 표처럼 내부 레이아웃이 중요한 요소는 개행 구조를 원형 그대로 보존하고, 페이지 경계에는 Form Feed 문자(0x0C)를 삽입하여 점자정보단말기에서의 페이지 구분을 보장한다.

2. Implementation of Semojum

세모점 구현을 위한 소프트웨어 및 프레임워크는 Table 3과 같다.

Table 3. Development Language & Framework

Front end	React
Backend	FastAPI
OCR & Layout	DeepSeek OCR 2
Image Captioning	GPT 4o, GPT-4.1 mini

1) Main page

세모점 서비스 애플리케이션의 메인 화면은 Fig. 5와 같다.

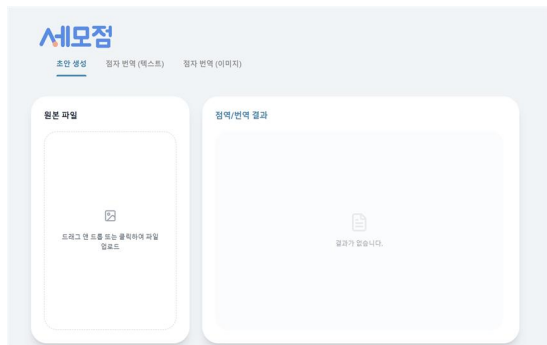


Fig. 5. Image captioning 7-pattern classification prompt

Fig. 5의 메인 화면 상단에는 서비스 로고와 함께 세 가지 처리 모드를 전환할 수 있는 탭 내비게이션을 배치한다. 세 가지 탭은 각각 초안 생성, 점자 번역(텍스트), 점자 번역(이미지)이며, 선택한 탭에 따라 하단 작업 영역의 입력 방식과 처리 파이프라인이 동적으로 전환된다. 하단 작업 영역은 좌측 원본 파일 입력 패널과 우측 점역/번역 결과 출력 패널로 이원화되어, 점역사가 원문과 점역 결과를 한 화면에서 동시에 비교하며 검수할 수 있도록 구성한다.

2) Draft Generation

초안 생성 모드 화면은 Fig. 6과 같다.



Fig. 6. Image captioning 7-pattern classification prompt

Fig. 6의 초안 생성 모드는 PDF, DOCX, HWP 등 복합 문서 전체를 입력받아 점역 초안을 자동으로 생성하는 기능을 제공한다. 좌측 '원본 파일' 패널은 드래그 앤 드롭 또는 클릭을 통해 파일을 업로드할 수 있는 입력 영역으로 구성한다. 파일이 업로드되면 BE의 비동기 병렬 파이프라인이 호출되어 OCR 모듈, 레이아웃 분석 모듈, 이미지 캡처링 모듈, 점역 엔진, 점자 규정 기반 레이아웃 제어 모듈이 순차적으로 실행된다. 각 모듈의 처리 결과는 FE로 스트리밍되어 우측 '점역/번역 결과' 패널에 순차적으로 표시된다. 이러한 파이프라인 구조는 장문의 전문 서적과 같이 처리 시간이 긴 문서에 대해서도 사용자에게 실시간 피드백을 제공한다.

3) Braille Translation (Text)

텍스트 기반 점자 번역 모드는 Fig. 7과 같다.



Fig. 7. Braille translation page (text)

텍스트 번역 모드는 사용자가 입력한 순수 텍스트를 한국 점자 규정에 따라 실시간으로 점역하여 출력하는 기능을 제공한다. 좌측 패널에 입력된 텍스트는 OCR·레이아웃 분석 과정을 거치지 않고 점역 엔진에 직접 전달되므로, 짧은 문장이나 제목·목차 등 구조가 단순한 입력에 대해 매우 빠른 응답 속도를 보장한다. 점역 결과는 우측 패널에 점자로 표시되며, 사용자는 결과를 직접 복사하거나 파일로 내보내 후속 작업에 활용할 수 있다.

4) Braille Translation (Image)

이미지 기반 점자 번역 모드는 Fig 8과 같다.



Fig. 8. Braille translation page (image)

이미지 번역 모드는 앞서 기술한 초안 생성 모드와 텍스트 점자 번역 모드를 하나의 파이프라인으로 통합한 end-to-end 모드이다. 이미지 형태의 복합 문서가 입력되면, 초안 생성 모드에서 사용되는 OCR·레이아웃 분석·이미지 캡처링·물리적 레이아웃 제어 모듈이 먼저 실행되어 문서의 구조와 텍스트를 추출한다. 이어서 텍스트 점자 번역 모드의 점역 엔진이 호출되어, 추출된 텍스트를 한국 점자 규정에 따라 점자로 변환한다. 최종 결과는 우측 패

널에 원문의 구조를 유지한 점자 형태로 표시된다.

V. Conclusions

본 논문은 기존 텍스트 중심 점역 도구의 한계를 극복하고, 수식·표·이미지가 혼재된 이공계 교재를 레이아웃 분석부터 BRF 출력까지 단일 파이프라인으로 처리하는 멀티모달 AI 점역 지원 서비스 세모점을 설계하고 구현하였다. 세모점의 핵심은 기존 도구들이 개별적으로 처리하거나 지원하지 않았던 레이아웃 감지, 시각 자료(표·이미지) 처리, 점역을 단일 워크플로우로 통합한 것이다.

DeepSeek-OCR-2 기반 레이아웃 분석, Semantic NMS, 표 레이아웃 최적화 엔진, GPT-4o 기반 이미지 캡처닝, Brailify와 커스텀 LaTeX 파서를 결합한 하이브리드 점역 엔진을 통해 이공계 교재의 다양한 문서 요소를 일관된 프로세스로 처리한다. 탭 기반 3모드 인터페이스와 Human-in-the-loop 구조는 점역사가 AI 초안을 문서 유형과 작업 단계에 따라 선택적으로 활용하고 검수할 수 있도록 지원한다. 현재 시스템의 주요 한계는 레이아웃 인식과 이미지 캡처닝의 성능 편차이다. 전문 점역사 검토를 통해 이미지 캡처닝 결과가 실용적 수준에 도달하였음을 확인하였다. 하지만 문서 유형과 내용에 따라 품질 편차가 존재하며, 수식 점역 커버리지도 사용 빈도가 낮은 일부 단일 기호 처리 부분에서 불완전한 부분이 남아 있다.

향후 연구에서는 점역사 사용자 테스트를 통한 작업 시간 단축 효과의 정량 측정, 레이아웃 인식 및 캡처닝 모듈의 체계적 성능 평가와 고도화, 그리고 한국 수학 점자 변환 엔진의 독립 오픈소스 모듈 공개이다.

ACKNOWLEDGEMENT

This work was supported by the SKTelecom FLY AI Challengers 8th program, conducted in collaboration with the Ministry of Employment and Labor and the Korean Skills Quality Authority as part of the 2025 K-Digital Training program.

REFERENCES

- [1] National Library for the Disabled, "Statistics on Braille Material Production," National Library for the Disabled, 2024.
- [2] Ministry of Education, "2025 Special Education Statistics," Special Education Policy Division, Ministry of Education, 2025.
- [3] Korea Blind Union, "Report on the Status of Braille Translators and Proofreaders," Korea Blind Union, 2024.
- [4] SENSEE, "Introduction to SENSEE Braille Editing and Automatic Conversion Service," sensimedia.co.kr, 2024.
- [5] Dot Incorporation, "Introduction to the Dot Pad Tactile Display Device," dotincorp.com, 2024.
- [6] Z. Zhao et al., "DocLayout-YOLO: Enhancing Document Layout Analysis through Diverse Synthetic Data and Global-to-Local Adaptive Perception," arXiv:2410.12628, Oct. 2024.
- [7] Y. Huang et al., "LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking," Proc. ACM Multimedia, pp. 4083-4091, 2022.
- [8] DeepSeek-AI, "DeepSeek-OCR-2," HuggingFace Model Hub, 2024.
- [9] J. Egli et al., "Liblouis: Open Source Braille Translator and Back-Translator," liblouis.io, 2024.
- [10] Ministry of Culture, Sports and Tourism, "Korean Braille Regulations (2024 Revision)," Ministry of Culture, Sports and Tourism Notice No. 2024-51, 2024.
- [11] OpenAI, "GPT-4o System Card," OpenAI, 2024.
- [12] National Library for the Disabled, "Guidelines for Braille Material Production (2025 Revision)," National Library for the Disabled, 2025.

Authors



Taemin Kim will receive the B.S. degrees in Computer Science and Electrical Engineering (double major) from the Korea Advanced Institute of Science and Technology (KAIST), Korea, in 2026. His research interests include natural language processing, large language models, and deep learning systems.



Junhyeok Lee will receive the B.S. degree in Computer Science and Engineering from Konkuk University, Korea, in 2026. His research interests include Android application development, front-end engineering, user interface design, and mobile software systems.



Haeun Cho will receive the B.S. degree in Computer Science from Dongduk Women's University, Korea, in 2026. Her research interests include backend development, server architecture, database systems, and scalable web services.



Hyunju Kim received the B.S. degree from the School of Robotics, Kwangwoon University, Seoul, Korea, in 2026. Her interests include computer vision, visual perception, AI-based image understanding, and AI service planning.



Sanghyun Kim will receive the B.S. degree from the school of electronic engineering, hanyang university, seoul, korea, in 2028. His research interests include wireless network technology and artificial intelligence.



Jimin Lee will receive the B.S. degree in Department of AI, Big Data & Management from Kookmin University, Korea, in 2027. Her research interests include large language models.



Won Joo Lee received the B.S., M.S. and Ph.D. degrees in Computer Science and Engineering from Hanyang University, Korea, in 1989, 1991 and 2004, respectively. Dr. Lee joined the faculty of the Department of

Computer Science and Engineering at Inha Technical College, Incheon, Korea, in 2008, where he has served as the Director of the Department of Computer Science and Engineering. He is currently a Professor in the Department of Computer Science and Engineering, Inha Technical College. He has also served as the president of The Korean Society of Computer Information. He is interested in parallel computing, internet and mobile computing, and cloud computing, data science, artificial intelligence.