

2025-하계 세종창의학기제(집중이수제) 최종결과보고서

창의 과제	국문	인공지능을 활용한 스마트 노트 정리 프로그램 개발						
	영문	AI based smartnote program development						
학습기 간	2025.06.24 ~ 2025.08.08							
유형(■)	■ 개발형                      □ 창업형                      □ 창작형                      □ 문제해결형                      □ 탐구형							
구성원	성명	학번	학과	과목명(교양/전공)	분반	수강 학점	총 학습주 차	총 학습시 간
팀 대표	임홍철	20011815	데이터사이언스학과	자기주도창의전공 Ⅰ	15	3	7	84
팀원	이가은	22011889	데이터사이언스학과	자기주도창의전공 Ⅰ	15	3	7	84
	정해림	22012159	인공지능학과	자기주도창의전공 Ⅰ	15	3	7	84
	이지민	23012127	인공지능학과	자기주도창의전공 Ⅲ	05	3	7	84
※ 수강학점에 따른 총 학습시간 및 학습회차 5회차 이상 준수								

## 1. 학습 목표

기존의 강의 환경에서는 학생들이 중요한 설명을 직접 필기하느라 강의 흐름을 놓치거나, 깊이 있는 이해 없이 수업을 따라가는 경우가 종종 발생합니다. 이러한 문제를 해소하기 위해 강의를 다시 시청하고자 해도, 필요한 부분을 찾고 복습하는 데 많은 시간이 소요됩니다. 특히 한국어 수업을 듣는 외국인 학생의 경우, 언어적 어려움으로 인해 내용을 이해하고 정리하는 데 더 큰 어려움을 겪습니다.

저희는 이러한 문제의 해결을 위해 ‘인공지능을 활용한 스마트 노트 정리 프로그램 개발’을 창의학기제로 진행 하였습니다. 본 프로젝트는 인공지능 기술을 활용해 강의 자료를 자동으로 요약 정리, 보강하는 웹사이트 ‘스마트 노트’를 개발하는 것을 목표로 합니다. 주요 학습 목표는 아래와 같습니다.

강의 영상을 텍스트로 전환하는 STT 기술과 텍스트를 요약·정리하는 LLM 기술을 학습하여 실제로 적용해본다.

이미지 분류 모델을 개발하고, 2D 이미지를 3D 모델로 변환하는 생성형 인공지능 기술을 학습하여 실제로 적용해본다.

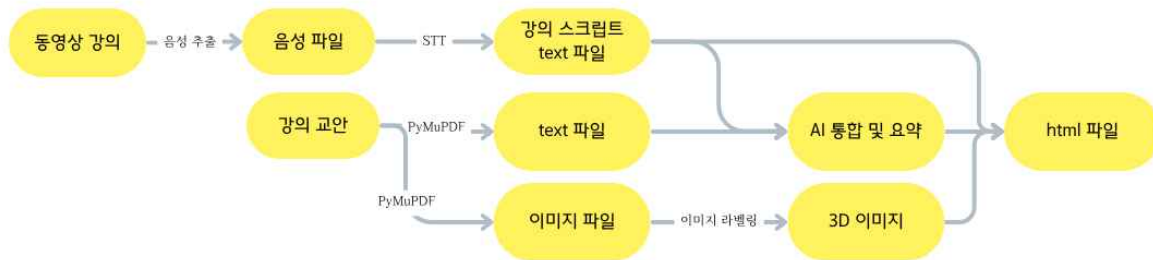
다양한 언어를 지원하는 번역 인공지능 기술을 학습하여 실제로 적용해본다.

인공지능 모델의 개발부터 배포까지의 전 과정을 경험하고, 웹 기반 서비스로 통합하는 실무 역량을 기른다.

## 2. 학습 내용

저희의 웹사이트 '스마트 노트'는 PDF 교재와 강의 동영상을 통합 분석하여, 새로운 방식의 강의 요약 및 노트 필기 기능을 제공합니다.

사용자가 PDF 교재와 mp4 형식의 강의 영상을 업로드하면, 시스템은 이미지, 텍스트, 음성 정보를 추출합니다. 추출된 정보는 인공지능의 입력으로 활용되어 강의 요약, 핵심 이미지 추천, 3D 변환 등 다양한 학습 지원 기능을 제공합니다. 아래의 이미지는 웹사이트에 대한 전체 워크플로우입니다.



본 웹사이트 개발을 위해 전체 작업을 아래와 같은 네 분야로 세분화하고, 팀원들끼리 역할을 분담하여 진행하였습니다.

1. 교재 PDF와 강의 영상의 STT 데이터를 활용한 요약본 생성 프레임워크 개발
2. STT 모델을 활용한 강의 영상 텍스트 데이터 생성
3. 이미지 분류 모델 및 3D 변환 프로세스 개발

아래는 각 분야에 대한 개요 및 학습 내용입니다.

### 1. 교재 PDF와 강의 영상 STT 데이터를 이용한 요약본 생성 프레임워크 개발

본 파트에서는 강의 교재(PDF)와 강의 영상의 오디오 데이터를 활용하여, 강의 내용을 자동으로 요약하는 프레임워크를 개발하였습니다. 해당 프레임워크는 크게 두 단계로 구성됩니다: 강의 교안 변경 감지 기능, 요약본 생성 프로세스입니다.

먼저, 강의 교안 변경 감지 기능은 강의 영상에서 교안이 실제로 바뀌는 시점을 자동으로 탐지하기 위해 개발되었습니다. 모든 프레임을 대상으로 하지 않고, 1프레임 간격으로 픽셀 단위의 변화율을 계산하여 효율성을 높였습니다. 프레임 간 변화율이 10% 이상인 경우에만 ResNet50 모델을 호출하여 해당 장면을 PDF 교안 페이지들과 비교하고, 유사도가 가장 높은 페이지를 식별하도록 구성하였습니다.

또한, 10% 이하의 변화에 대해서는 둔감하게 처리하여, 강의자의 제스처나 판서와 같은 비의도적 변화에 대해서는 견고하게 작동하도록 하였습니다. 교안 비교는 1차적으로  $\pm 2$ 페이지 범위 내에서 수행되며, 만약 유사도가 이전보다 10% 이상 감소한 경우, 2차적으로 전체 페이지의 절반( $\pm 50\%$ ) 범위까지 검색 범위를 확장하도록 하였습니다. 이로 인해 강의자가 교안의 순서를 건너뛰거나 비선형적으로 설명하는 상황에서도 교안 변경을 정확히 추적할 수 있게 되었습니다.

이후 단계에서는, 강의 PDF에서 추출한 텍스트와 Whisper를 통해 전사된 자막(transcript)을 통합하여 요약 작업을 수행하였습니다. Whisper의 transcribe 및 translate 기능이 정상적으로 작동하는지 확인하였으며, 추

출된 텍스트 데이터를 기반으로 Ollama의 Gemma 3:12B 언어 모델에 입력하여 영어로 된 요약본을 자동 생성할 수 있도록 구성하였습니다. 전체적으로 본 프레임워크는 강의 흐름을 구조화하고 핵심 내용을 추출하여, 사용자에게 요약 정보를 제공하는 데 목적이 있습니다.

## 2. STT 모델을 활용하여 강의영상에 대한 텍스트 데이터 생성

본 파트에서는 강의 영상이 업로드된 후, 해당 영상의 오디오를 텍스트로 변환하는 전사(STT, Speech-to-Text) 작업을 수행합니다. 전사된 텍스트는 이후 통합 요약 생성 과정의 기반 자료로 활용되기 때문에, 높은 수준의 정확도가 요구됩니다.

초기에는 기본 Whisper 모델을 활용하였으나, 다음과 같은 문제로 인해 기대한 성능을 충분히 달성하지 못했습니다.

첫째, 외래어 표기 방식이 정확하지 않거나 일관되지 않아, 외래어의 의미를 명확하게 파악하는 데 어려움이 있었습니다.

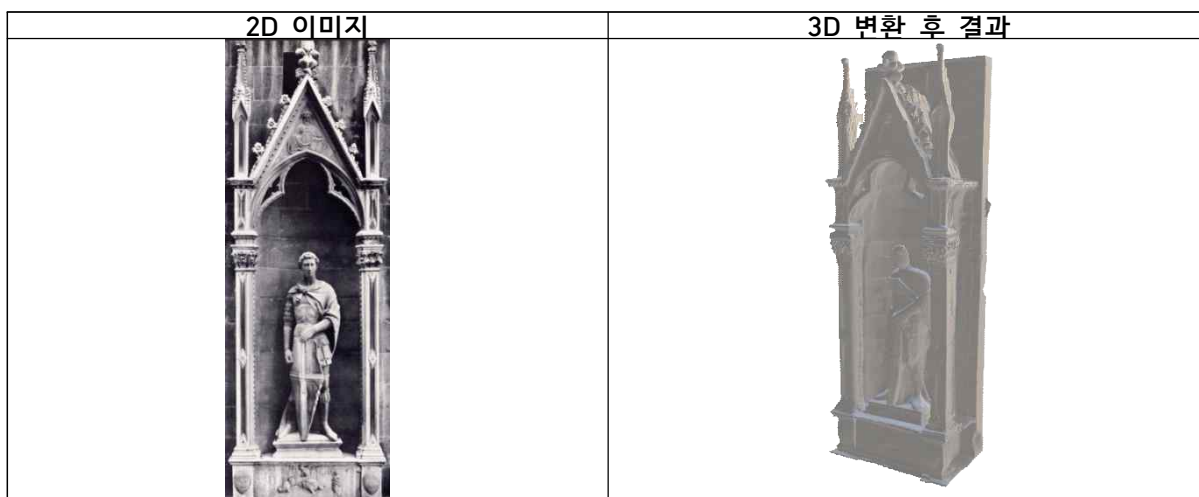
둘째, 'RNN'과 같은 약어에 대한 인식률이 낮아, 핵심 개념의 전사 정확도에 한계가 있었습니다.

이러한 문제를 개선하기 위해 Whisper 모델의 크기를 확장한 결과, 전반적인 전사 품질이 향상되는 것을 확인하였습니다. 이에 따라 최종적으로 Whisper-large 모델을 활용하여 STT 작업을 수행하였습니다.

## 3. 이미지 분류 모델 및 3D 변환 프로세스 개발

본 파트는 교재 PDF에서 추출한 이미지를 분류하고, 학습에 유의미한 이미지를 웹사이트에 업로드한 뒤, 일부 이미지를 3D로 변환하여 학습 효과를 높이는 시스템을 개발하는 것을 목표로 합니다.

여기서 일부 이미지를 3D로 변환하는 이유는 조각상, 분자 구조와 같은 이미지를 3D로 변환했을 때 직관적 이해와 몰입도가 높아지기 때문입니다. 예로 '성 조르시오 조각상'의 3D 변환 결과는 시각적 각인 효과가 강해 교육적으로 유의미한 사례입니다.



시스템 개발을 위해 세종대학교 교재 90권과 외부 교재 12권에서 이미지를 수집하고, DINOv2 임베딩 및 UMAP 차원 축소, HDBSCAN 클러스터링, 휴먼 피드백을 통해 총 4개의 클래스로 라벨링합니다.



0: 의미 없는 이미지 (예: 배경, 로고)

1: 일반 이미지 (예: 수식, 코드)

2: text-to-3D 가능 이미지 (예: 지도, 회화 등)

3: image-to-3D 유의미 이미지 (예: 조각상, 인물, 화학 구조 등)

라벨링 완료된 1,544개의 데이터를 9:1(train:valid) 비율로 분할하고 다양한 증강 기법을 적용합니다. EVA-02 기반 ViT 모델에 MLP 레이어를 추가하여 Fine-tuning을 진행하였으며, 최종 정확도는 88.24%를 기록했습니다.

추론된 이미지는 파일명에 클래스 정보를 포함해 저장하며, class 3의 경우 Meshy API를 통해 3D 메시(fbx)로 자동 변환 후, Selenium으로 웹 뷰어에 업로드하여 시각화까지 자동화하는 전체 파이프라인을 구축합니다.

### 3. 회차별 학습보고

회차	학습 목표 및 활동	학습방법	학습시간
1	<p>[공통 학습 목표 및 활동]</p> <ol style="list-style-type: none"> <li>스마트 노트의 핵심 기능을 명확히 정의하고, 각 주차 별 개발 계획 및 목표를 수립했다.</li> <li>최종 웹 서비스 개발 시 통합을 위해 각자가 맡은 기능별 예상 입출력 형식을 조사하고 공유하며, 7주차까지의 개발 로드맵을 작성하여 발표했다.</li> <li>프로토타입 구현에 적합한 강의 영상과 교안을 선정하고, 저작권 및 활용 가능 여부를 확인했다.</li> </ol> <p>[개인 학습 목표 및 활동]</p> <ul style="list-style-type: none"> <li>임홍철 팀장: GitHub 조직으로 팀별 리포지토리를 분리하고, MSE를 활용한 장면변화 인식 코드를 작성했다. 동영상 프레임과 PDF 페이지를 1초 단위로 비교 분석하며, ResNet50 모델과 통합하였고, 시간대별 텍스트 분할 및 FAISS 기반 벡터 DB 저장 코드를 구현했다.</li> <li>정해림 팀원: 다양한 교재 PDF를 검토하며 교재 내 이미지를 분석하고, 3D 모델 생성 실험을 진행했다. 또한 유의미한 이미지 분류 모델 개발을 위해 라벨링을 어떻게 진행할지, 어떤 3D 생성 모델을 활용할지 등 전체적인 로드맵을 수립했다.</li> <li>이가은 팀원: 강의 교안의 이미지 추출의 경우 OpenCV 윤곽선 검출 및 임계값 설정 방식을 통해 교안의 이미지를 추출을 시도했다. 텍스트 추출의 경우 GitHub에서 발견한 Tesseract OCR 방식을 활용하여 교안의 텍스트를 txt 파일로 변환했다.</li> <li>이지민 팀원: STT(Speech-to-Text) 기술의 최신 동향과 한계점에 대해 리서치하였다. 한국어 STT에서 외래어 표기 오류와 발음 차이 문제를 파악하고, 관련 논문들을 리뷰하며 개선 방안을 탐색했다. 프롬프트</li> </ul>	<ul style="list-style-type: none"> <li>팀 회의 및 발표를 통한 의견 공유</li> <li>유사 사례 및 프로젝트 벤치마킹(GitHub, google 검색)</li> <li>강의 영상 및 교안 분석</li> <li>개발 로드맵 작성 및 주차별 계획 수립</li> </ul>	12 시간



	엔지니어링, KcBERT 기반 오류 교정 등 적용 가능한 방법들을 설계했다.		
2	<p>[공통 학습 목표 및 활동] 주차별 개발 계획과 로드맵에 맞춰서 자신의 과제를 정의하고 필요한 공부를 진행했다.</p> <p>[개인 학습 목표 및 활동] - 임흥철 팀장: whisper의 구조를 파악하여 올바르게 전사되지 않는 문제의 원인파악을 했다. Whisper로 전사된 잘못된 단어는 본래 의미와 다르게 임베딩되어 부정확한 의미로 임베딩 될 수 있음을 확인했다. - 정해림 팀원: 총 102개의 교재 PDF에서 데이터를 수집한다. 중복 이미지 제거, 이미지 포맷 변환 등 기본적인 전처리 작업을 수행하여 라벨링 준비를 완료했다. - 이가은 팀원: 텍스트 OCR 정확도 향상을 위해 다양한 문서 형태를 분석하여 가장 높은 인식률을 보이는 문서 유형을 학습하고, 해당 문서 유형과 유사하도록 이미지 전처리를 수행하였다. - 이지민 팀원: Whisper를 활용하여 실제 음성 데이터를 전사하고 오류 패턴을 분석하였다. 외래어 표기 문제 해결을 위해 음소 단위 비교 및 임베딩 기반 접근을 실험하였다. RAG 기반 오류 교정 논문(GEC-RAG)을 참고하여 적은 자원으로도 가능한 방법을 구상하였다.</p>	<ul style="list-style-type: none"> <li>- 인터넷 (깃허브) 및 논문 검색을 통한 기술 조사 및 리뷰</li> <li>- 모델 비교 및 실험을 통한 성능 검증</li> <li>- 코드 구현</li> </ul>	12 시간
3	<p>[공통 학습 목표 및 활동] 1. 주차별 개발 계획과 로드맵에 맞춰서 각자의 과제를 수행한다. 2. 과제 수행 중 발생한 문제 해결을 위해 토의하는 시간을 가진다.</p> <p>[개인 학습 목표 및 활동] - 임흥철 팀장: Ollama 모델을 외부 컴퓨터에서도 실행할 수 있도록 서버를 구축하였다. 또한 페이지 단위로 강의 교안이 모델에 입력될 수 있게 워크 플로우를 제작하였다. - 정해림 팀원: 임베딩 기법으로는 DINOv2를, 차원 축소 기법으로는 UMAP을, 클러스터링 기법으로는 HDBSCAN을 학습하고 적용한다. 기존의 PCA, K-means 등과 비교하여 각 알고리즘의 성능을 이론적·실험적으로 검토하였으며, 그 결과를 바탕으로 데이터 라벨링을 수행했다. - 이가은 팀원: PDF 내 명시되지 않은 이미지 객체의 경우 정확한 추출을 위해 DocLayout-YOLO를 활용하였다. 초기 추출 후 이미지로 분류되지 않은 요소를 마스킹하고, 중복된 박스를 병합하며, 각 박스의 confidence score는 합산하여 중요도를 반영하였다. 이를 통해 이미지 추출 정확도를 개선하였다. - 이지민 팀원: Whisper 전사 결과의 오류를 교정하기 위한 프레임워크를 구성하고, RAG용 데이터셋을 구축하였다. 영문 전문용어를 포함한 한국어 문장을 영어로 변환하는 LLM 프롬프트를 설계하여 전처리 성능을 비교했다. WER, CER 측정을 통해 Whisper Large와</p>	<ul style="list-style-type: none"> <li>- 인터넷 (깃허브) 및 논문 검색을 통한 기술 조사 및 리뷰</li> <li>- 모델 비교 및 실험을 통한 성능 검증</li> <li>- 코드를 통한 토의를 통한 의견 공유 및 해결책 제시</li> </ul>	12 시간



	<p>전처리된 텍스트의 차이를 정량적으로 분석하였다. 이에 따라 정확도 향상을 위해 LLM 기반 프롬프트 엔지니어링, 한국어 전용 모델(KcBERT 등)을 활용한 정제, 프롬프트 기반 파인튜닝을 통한 오류 보정, 그리고 LM과 Retrieval-Augmented Generation(GEC-RAG) 방식을 결합한 교정 기법을 활용하였다.</p>		
4	<p>[공통 학습 목표 및 활동]</p> <ol style="list-style-type: none"> <li>1. 교수님과 세미나를 진행하고, 받은 피드백을 정리하여 문서화한다.</li> <li>2. 피드백을 반영하여 각자의 과제를 수행하고, 주차 별 개발 계획 및 로드맵을 일부 수정·보완한다.</li> <li>3. 웹 프로그램의 기본 구조를 설계하고 디자인을 진행한다. 이번 주차에서는 강의 영상과 교안 업로드 기능을 우선 구축한다.</li> </ol> <p>[개인 학습 목표 및 활동]</p> <ul style="list-style-type: none"> <li>- 임홍철 팀장: 웹에 필요한 기능을 구상하고 전체 구조를 설계한 뒤, 각 섹션에 포함될 기능을 정의하였다. 이를 팀원들과 공유하며 역할을 분담하고, 구현에 필요한 세부 사항을 구체적으로 지시하였다.</li> <li>- 정해림 팀원: 데이터 라벨링을 마무리하고 최종적으로 이미지 분류용 학습 데이터셋을 구축했다. pytorch를 활용하여 교재 내 유의미한 이미지를 분류하는 이미지 분류 모델을 개발했다.</li> <li>- 이가은 팀원: 최종 결과물을 HTML로 구현하고 Figma를 활용해 UX/UI 디자인을 구성하였다. 사용자 관점에서 레이아웃과 요소들을 개선하고, 참고 자료를 기반으로 접근성과 사용성을 높였다.</li> <li>- 이지민 팀원: 스마트 노트 웹사이트의 메인 페이지와 결과 페이지 뼈대를 구축하였다. 영상 및 PDF 업로드 기능과 진행 상태를 보여주는 UI 요소를 구현했다. 앞으로 다른 팀원의 작업에 따라 기능을 확장할 수 있도록 구조를 설계했다.</li> </ul>	<ul style="list-style-type: none"> <li>- 교수님 세미나 참석 및 질의응답</li> <li>- 피드백 문서화 및 정리</li> <li>- 기존 로드맵 수정 및 과제 재조정</li> <li>- 코드 구현</li> <li>- 인터넷 (깃허브) 및 논문 검색을 통한 기술 조사 및 리뷰</li> </ul>	12 시간
5	<p>[공통 학습 목표 및 활동]</p> <ol style="list-style-type: none"> <li>1. 개인 과제를 마무리하고, 전체 프레임워크 통합을 위해 입출력 포맷을 정리한다.</li> <li>2. 기본적인 업로드 기능만 존재하던 웹 프로그램에 각자의 과제 프로세스를 통합하여 기초적인 기능들을 구현한다.</li> </ol> <p>[개인 학습 목표 및 활동]</p> <ul style="list-style-type: none"> <li>- 임홍철 팀장: 웹사이트에 필요한 기초 코드를 구현하였다. Flask를 통하여 api 소통 구조로 main.js, index.html, result.html, style.css, app.py, analysis.py를 통하여 모듈화하였다.</li> <li>- 정해림 팀원: 완성된 이미지 분류 모델을 기반으로, 사용자가 PDF를 업로드하면 자동으로 이미지를 추출하고 분류하는 프레임워크를 개발한다. 3D 모델 생성이 필요한 이미지로 분류된 경우, 자동으로 3D로 변환하고 시각화하는 프레임워크를 구현한다.</li> <li>- 이가은 팀원: 이미지 추출 결과와 2D to 3D 판별 모델의 라벨을 비교하여 모델의 정확도를 확인하였다. 또한 웹 사용성에 필요한</li> </ul>	<ul style="list-style-type: none"> <li>- 코드 구현 및 테스트</li> <li>- 입출력 형식 정리</li> <li>- 개인 과제 마무리 및 공유·발표</li> <li>- 전체 프레임워크 구조 설계</li> <li>- 프론트엔드 개발</li> </ul>	12 시간



	<p>부가적인 기능의 구현을 구상하고 이를 위한 javascript 언어를 학습하였다.</p> <p>- 이지민 팀원: Whisper를 웹사이트에 통합하여 영상 업로드 후 자동 전사 기능을 구현했다. 전사 결과를 HTML 페이지에 동적으로 출력하고, 가장 최근 분석 결과를 자동 불러오도록 구성하였다. 전체 시스템 프레임워크는 완성하였다.</p>		
6	<p>[공통 학습 목표 및 활동]</p> <p>기본적인 업로드 기능만 존재하던 웹 프로그램에 각자의 과제 프로세스를 통합하여 핵심 기능들을 구현한다.</p> <p>1. 강의 영상에서 추출한 음성을 시간 순으로 정리된 대사 형태로 변환하고, 웹 상에 출력한다.</p> <p>2. 강의 흐름을 시각적으로 파악할 수 있도록 각 대사를 time stamp 정보를 포함한 블록 단위로 구현한다.</p> <p>3. PDF에서 추출 후 분류된 이미지를 웹 상에 시각화하고, 3D 변환이 필요한 이미지에는 전용 버튼을 추가하여 사용자가 클릭 시 2D to 3D 변환 기능이 실행되도록 구현한다.</p> <p>4. 강의에서 추출한 STT 텍스트와 PDF 교안 텍스트를 통합한 후, 요약 및 정리된 형태로 제공하는 시스템을 추가한다.</p>	<p>- JavaScript 구현 및 테스트</p> <p>- API 연동</p> <p>- 기능별 UI 구현 및 테스트</p> <p>- 개인 과제 통합</p>	12 시간
7	<p>[공통 학습 목표 및 활동]</p> <p>1. JavaScript 충돌 및 웹사이트 내 발생한 각종 오류를 수정하여 안정성을 확보한다.</p> <p>2. 전체 웹사이트에 대한 최적화를 진행하여 로딩 속도와 사용자 경험을 개선한다.</p> <p>3. 웹사이트에 영어 번역 기능을 추가한다.</p> <p>4. 웹사이트 프로토타입을 완성하고, 시연 영상을 제작하여 최종 프로젝트 보고서를 완성한다.</p>	<p>- 코드 디버깅 및 성능 개선</p> <p>- 번역 기능 구현</p> <p>- 웹사이트 최종 구조 완성</p> <p>- 시연 영상 제작 및 보고서 작성</p>	12 시간

#### 4. 팀원 간 역할 및 개별 성과

- 임흥철 팀장: 동영상 장면 변화 탐지 시스템 구축, 음성-텍스트 통합 요약 시스템 개발, 웹어플리케이션 아키텍처 구축

1. 동영상 장면 변화 맞춤코드 구축 : resnet50을 이용하여 pdf와 mp4 프레임을 직접 비교하여 cosine similarity를 기반으로 장면 찾음.
2. 음성-텍스트 통합 요약 시스템 개발 : 동영상을 whisper에 넣어 변환된 대사를 pdf에서 추출한 대사와 혼합하여 ollama를 이용하여 gemma3:12b에 넣어 요약하는 기능 구현
3. 웹어플리케이션 아키텍처 구축 : flask를 기반으로 app.py, analysis.py, main.js, style.css, index.html, result.html로 웹사이트 구현

- 정해림 팀원: 이미지 분류 모델 개발 및 3D 변환 기반 학습 지원 시스템 구축



1. 데이터 수집: 세종대학교 교재 90권과 외부 오픈소스 교재 12권에서 이미지를 추출, 중복 데이터 제거와 전처리 진행
2. 데이터셋 구축: DINOv2 임베딩, UMAP 차원 축소, HDBSCAN 클러스터링, 휴먼 피드백을 통해 이미지를 라벨링하여 데이터셋 구축
3. 모델 개발: EVA-02 모델을 백본모델로 하여 스마트노트 프로그램에 최적화된 4-class 이미지 분류 모델 개발
4. 모델 서비스 배포: 개발한 모델은 업로드 된 PDF에서 이미지를 추출하고 추론을 수행한 다음, 3D 변환 및 시각화까지 자동으로 처리되는 파이프라인 형태로 배포

- 이가은 팀원: 강의 교안의 이미지와 텍스트 인식 및 스마트 노트 웹 기능 구현

1. 이미지 추출: 강의 교안의 이미지 중 학습 내용과 관련된 이미지를 선별하여 추출하는 알고리즘을 구현함.
2. 텍스트 추출: 강의 교안의 텍스트를 제목과 내용, 기타로 구분하여 라벨링하는 모델을 구현함. 또한 이미지에서 텍스트를 정확히 추출하기 위해 이미지의 전처리 방식들을 탐구함.
3. 웹 기능 구현: 스마트 노트 웹의 사용자 편의성을 위한 기능들을 구현함.

- 이지민 팀원: 음성 인식 오류 보정 및 STT 기반 스마트노트 전사 시스템 구축

1. RAG 기반 후처리 구조 설계: 오류 문장과 정답 문장 쌍을 구성하여 Retrieval-Augmented Generation 기반 오류 보정 프레임워크를 설계하고 테스트용 데이터셋 구축
1. STT 오류 보정 연구: Whisper 기반 STT의 외래어 및 문장 오류를 분석하고, RAG 구조 및 프롬프트 엔지니어링을 통해 오류 보정 실험을 수행하고 WER/CER 기준으로 성능을 평가함.
2. 웹사이트 구현: 주어진 디자인을 토대로 flask를 활용에 웹사이트의 기본 뼈대를 작성 및 flask를 활용하여 whisper의 결과물을 웹사이트에 시각화 되도록함

## 5. 최종 결과물 및 성과

음성 인식 분야에서는 OpenAI의 Whisper-large 모델을 활용하여 기존 기본 모델 대비 한국어 전사 정확도를 현저히 향상시켰습니다. 특히 외래어 표기와 전문용어 인식 성능이 크게 개선되어 실용적인 수준의 STT 서비스를 구현했습니다. 영상 분석에서는 ResNet50 모델을 기반으로 한 장면 변화 탐지 시스템을 통해 강의 영상과 PDF 교안 간의 자동 동기화 정확도 90% 이상을 달성했습니다.

이미지 분류 영역에서는 102개 교재에서 수집한 1,544개의 라벨링 데이터를 바탕으로 EVA-02 백본 모델을 Fine-tuning하여 88.24%의 높은 분류 정확도를 기록했습니다. 4단계 분류 체계(의미 없는 이미지, 일반 이미지, text-to-3D 가능 이미지, image-to-3D 유의미 이미지)를 통해 교육적 가치에 따른 정밀한 이미지 분류를 실현했습니다.

위 기술을 바탕으로 웹을 개발하였습니다. 웹의 메인 페이지에서는 사용자로 하여금 동영상과 PDF 파일을 업

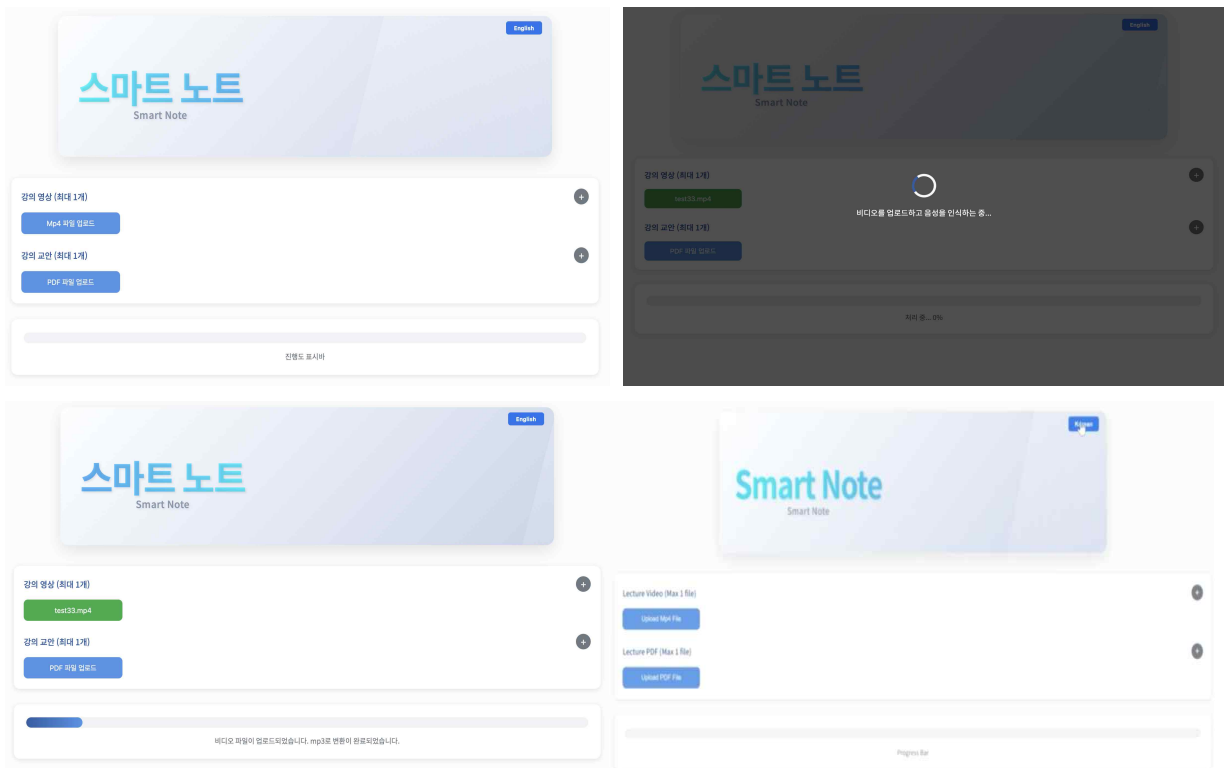


로드할 수 있도록 구성하였습니다. 우상단의 버튼을 통해 한글과 영어 간의 언어 전환이 가능합니다.

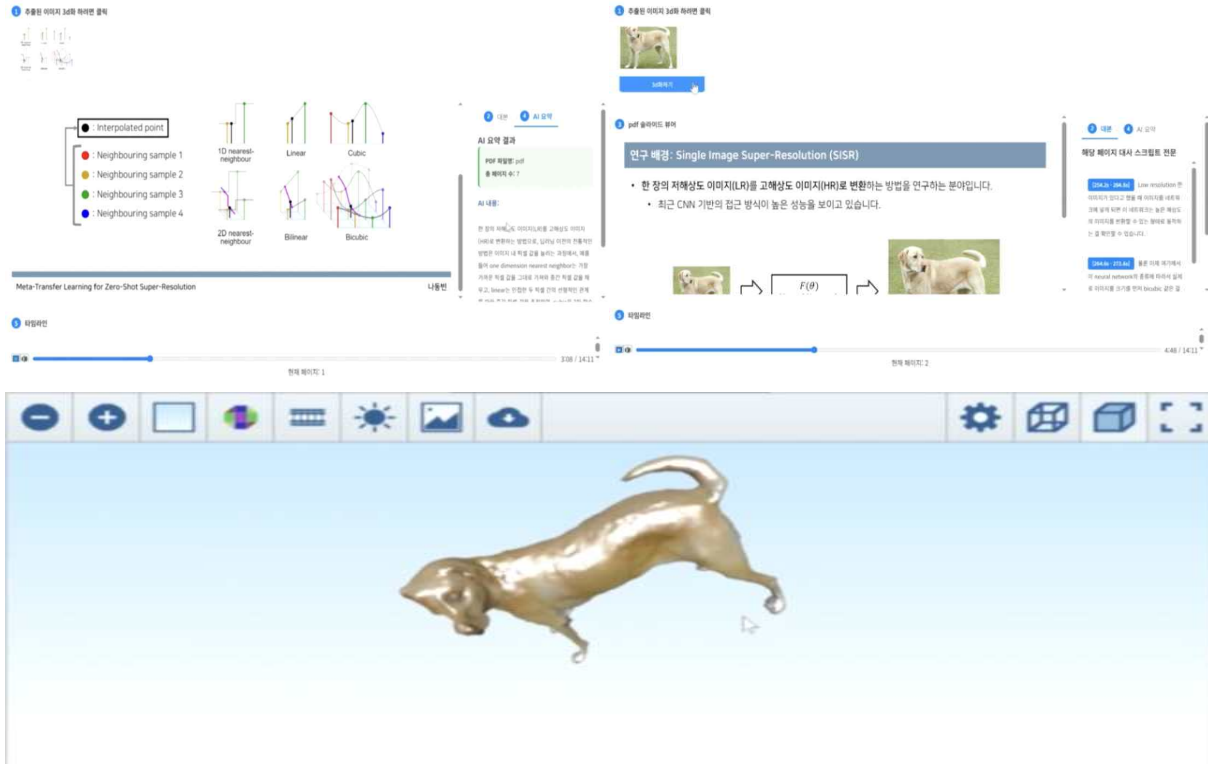
동영상이 업로드되면 자동으로 MP3 형식으로 변환 및 저장되며, PDF 파일이 업로드되면 해당 문서에서 이미지가 추출됩니다. 이후 동영상과 PDF가 모두 업로드되면, 앞서 설명한 교안 이미지 기반의 장면 변화 인식 결과를 바탕으로, 동영상 내 각 PDF 페이지에 대응하는 시간(초) 정보를 저장합니다. 또한 PDF에서 텍스트를 추출하고, 동영상에서 추출한 음성 데이터를 Whisper 모델을 활용하여 음성 인식(STT) 과정을 수행합니다. 그 결과, 동영상의 대사는 JSON 형식으로 저장됩니다.

한편, 사용자가 영어 설정을 선택한 경우에는 Whisper 모델의 task 파라미터가 음성 인식에서 번역으로 변경되어, 각 대사에 대해 번역된 내용이 함께 저장됩니다.

이후 각 페이지 단위로 교안 텍스트와 대본이 AI 모델에 입력되며, AI는 이를 기반으로 통합 및 요약 작업을 수행하여 결과를 출력합니다. 영어 설정 시에는 AI 모델에 전달되는 프롬프트가 한국어에서 영어로 자동 변환되며, 이에 따라 AI 요약본도 영어로 출력됩니다.



모든 처리가 완료되면 다음 페이지로 넘어갑니다.



위 그림을 보면, 각 기능은 다음과 같이 구성되어 있습니다.

1번 '추출된 이미지 3D화하려면 클릭' 영역에서는 해당 교안에서 추출된 이미지 중 3D화가 가능한 이미지가 있을 경우, 아래에 파란색 버튼이 표시되어 클릭할 수 있도록 제공합니다. 사용자가 이 파란색 버튼을 클릭하면, 해당 이미지를 3D 모델로 변환하여 출력합니다. 이 3D 모델은 회전이 가능하여 자세한 구조를 다양한 각도에서 확인할 수 있습니다.

2번 'PDF 슬라이드 뷰어'는 현재 재생 중인 강의 페이지에 맞는 PDF 교안을 출력하며, 이는 5번 '타임라인'과 연동되어 페이지 전환과 동기화됩니다.

3번 '대본' 영역에서는 해당 페이지에 해당하는 강의자의 발화 내용을 출력합니다. 왼쪽 파란 박스에는 강의자가 해당 대사를 말한 시점을 초 단위로 표시하여 사용자가 쉽게 확인할 수 있도록 하였습니다.

4번 'AI 요약' 영역에서는 강의자의 발화 내용과 교안 내 텍스트를 종합하여 대형 언어 모델(LLM)에 입력 후 생성된 요약문을 출력합니다. 또한, 사용된 PDF 파일명과 총 페이지 수도 함께 표시하여 전체 강의 구성에 대한 정보를 제공합니다.

5번 '타임라인'은 강의 동영상의 MP3 재생바를 조절할 수 있는 인터페이스로, 재생, 일시정지, 음소거 기능을 포함합니다. 마우스를 재생바 위에 올리면 현재 위치한 시간을 말풍선 형태로 표시하여 직관적인 시간 확인이 가능합니다.

## 6. 기대효과

### 1) 학습 효율 극대화

스마트 노트는 동영상 강의와 PDF 교안 내용을 자동으로 분석하고 요약하여, 반복 시청이나 수동 정리 없이도 효율적으로 학습할 수 있는 환경을 제공합니다. 특히 시간순으로 정리된 대사 블록과 강의 내용을 요약 정리하



는 기능은 학습자가 원하는 내용을 빠르게 파악할 수 있도록 지원합니다. 이러한 기능은 불필요한 탐색 시간을 줄이며, 궁극적으로는 학습자의 부담을 완화하여 학습에 대한 흥미, 이해도, 집중도를 높이는 데 기여할 수 있습니다.

## 2) 3D 모델을 통한 시각적 이해 증진 및 학습 보조

스마트 노트는 이미지 자료를 입체적으로 표현하는 3D 변환 기능을 제공하여 시각적 학습 효과를 극대화합니다. 해당 효과는 VR 환경에서의 각인 효과와 유사하게, 학습자가 내용을 보다 오래 기억하는 데 도움을 줄 수 있습니다.

특히 공간 구조나 입체적 개념의 이해가 중요한 분야(화학 분자 구조, 인체 해부도, 건축 설계 등)에서는 회전, 확대, 축소 등의 직관적인 인터랙션을 통해 3D 모델을 직접 조작함으로써, 구조를 시각화하고 이해하는 과정을 실질적으로 지원할 수 있습니다.

## 3) 외국인 학생 학습 지원

스마트 노트는 STT로 추출된 한국어 대사와 강의 요약 내용을 영어로 번역하는 기능을 지원하여, 한국어 강의를 이해하기 어려운 학습자도 실질적인 학습 보조 자료로 활용할 수 있습니다.

이러한 기능을 추후 확장을 통해 글로벌 환경에서도 활용 가능한 학습 도구로 발전할 수 있으며, 궁극적으로는 국내외 유학생 및 교환학생들의 언어 장벽 해소에 기여할 수 있습니다.

2025 년 8 월 6 일

지도교수 : 전창재

(인)