JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 30 No. 3, pp. 21-30, March 2025
https://doi.org/10.9708/jksci.2025.30.03.021

A Design and Implementation CODA Service Application Based on Generative AI

Jeongmin Ahn*, Hyewon Ryu**, Dusan Baek***, Gyeongho Cho****,
Seongbin Choi*****, Ho-Young Kwak†, Won Joo Lee‡

*Student, Dept. of Statistics, Sookmyung Women's University, Seoul, Korea
**Student, Dept. of Industrial and Management Engineering, Korea University, Seoul, Korea
***Student, Dept. of Material Science and Engineering, Ulsan National Institute of Science and Technology, Ulsan, Korea
****Student, Dept. of Mechanical Engineering, Pusan National University, Pusan, Korea
*****Student, Dept. of Geography with a Double Major in Computer Engineering, Kyunghee University, Seoul, Korea
† Professor, Dept. of Computer Engineering, Jeju National University, Jeju, Korea
‡ Professor, Dept. of Computer Science & Engineering, Inha Technical College, Incheon, Korea

[Abstract]

In this paper, we proposes a CODA service application equipped with a sign language translation function based on generative AI. The application is characterized by its sign language translation technology and a user interface (UI) that minimizes text. The sign language translation technology consists of the Image2Pose, Pose2Gloss, Gloss2Text, and Image Generation modules. The Image2Pose module analyzes input sign language videos and converts them into images. The Pose2Gloss module translates the meaning of Gloss based on continuous Pose key points. The Gloss2Text module converts Gloss into text, while the Image Generation module converts text into images. The application undergoes field testing with actual CODA users and is refined based on their feedback. Additionally, the need for additional sign language features is identified, and these are implemented using Contrastive Learning. The proposed application is expected to contribute not only to smooth communication within CODA families but also to the advancement of sign language translation technology.

▸ **Key words**: Shared Album, Graph Convolution Network, Natural Language Processing, Generative AI

• First Author: Jeongmin Ahn, Corresponding Author: Ho-Young Kwak, Won Joo Lee
*Jeongmin Ahn (ajm1016@sookmyung.ac.kr), Dept. of Statistics, Sookmyung Women's University
**Hyewon Ryu (abc67432@korea.ac.kr), Dept. of Industrial and Management Engineering, Korea University
***Dusan Baek (santoo@unist.ac.kr), Dept. of Material Science and Engineering, Ulsan National Institute of Science and Technology
****Gyeongho Cho (gh_cho@pusan.ac.kr), Dept. of Mechanical Engineering, Pusan National University
*****Seongbin Choi (beani28@khu.ac.kr), Dept. of Geography with a Double Major in Computer Engineering, Kyunghee University
†Ho-Young Kwak (kwak@jejunu.ac.kr), Dept. of Computer Engineering, Jeju National University
‡Won Joo Lee (wonjoo2@inhatc.ac.kr), Dept. of Computer Science & Engineering, Inha Technical College
• Received: 2024. 12. 30, Revised: 2025. 01. 18, Accepted: 2025. 01. 20.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

[요 약]

본 논문에서는 생성형 AI 기반의 수어 번역 기능을 탑재한 CODA service 애플리케이션을 제안한다. 이 애플리케이션은 수어 번역 기술과 텍스트를 최소화한 사용자 인터페이스(UI)를 주요 특징으로 한다. 수어 번역 기술은 Image2Pose, Pose2Gloss, Gloss2Text, Image Generation 모듈로 구성한다. Image2Pose 모듈은 수어 영상을 입력받아 내용을 분석하고 이를 이미지로 변환하는 기능을 제공한다. Pose2Gloss 모듈은 연속적인 Pose 키 포인트를 입력받아 어떤 의미의 Gloss인지 번역하는 기능을 제공한다. Gloss2Text 모듈은 글로스를 텍스트로 변환하는 기능을 제공한다. Image Generation 모듈을 통해 텍스트를 이미지로 변환한다. 이 애플리케이션은 실제 사용자인 CODA를 대상으로 필드 테스트를 실시하고, 그 피드백을 바탕으로 수정 보완한다. 또한, 수어 추가 기능의 필요성을 확인하고, 이를 대조 학습(Contrastive Learning)을 활용하여 구현한다. 본 연구에서 제안하는 애플리케이션은 CODA 가족 간 원활한 소통을 지원할 뿐만 아니라, 수어 번역 기술 발전에도 기여할 수 있을 것으로 기대한다.

▸ **주제어**: Shared Album, Graph Convolution Network, Natural Language Processing, Generative AI

I. Introduction

CODA(Children of Deaf Adults)는 농인, 농인 부모와 자녀를 의미한다. 전 세계적으로 CODA 인구는 농인 인구에 비례하며, 농인 부모의 상당수가 CODA 자녀를 둔 것으로 알려져 있다[1]. 한국의 경우 약 30만 명의 농인이 있는 것으로 추산되지만, CODA에 대한 구체적인 통계는 현재 부족한 상황이다[2]. CODA는 가족 내에서는 수어를 주로 사용하지만, 학교나 사회생활에서는 구어와 텍스트 기반 소통을 병행하는 이중언어 환경에서 자란다. 일부 CODA는 수어를 충분히 배우지 못해 의사소통에서 어려움을 겪기도 하며, 이 경우 비언어적 신호나 텍스트 기반 소통 방법으로 문제를 극복하기도 한다.

최근 AI 기술의 발전으로 CODA들이 겪는 소통의 어려움을 해결하려는 서비스가 증가하고 있다. CODA의 주요 소통 방식으로 영상 통화나 음성-텍스트 변환 기술을 활용한 서비스가 개발되고 있다[3]. 현재의 수어 번역 시스템은 주로 신체 동작(Pose)을 텍스트로 변환하는 데 초점을 맞추고 있다[3]. 이러한 기술은 기본적인 소통은 지원하지만, 다양한 소통 방식을 포함하지는 못하는 단점이 있다. CODA 가족 내에서 수어를 사용하는 사람과 사용하지 못하는 사람이 공존하기 때문에, 수어뿐만 아니라 텍스트와 음성 인식까지 지원하는 기능이 필요하다. 현재의 수어 번역 기술은 글로스(Gloss) 기반의 의사소통을 충분히 반영하지 못하고 있다. 글로스는 수어에서 의미적 최소 단위를 나타내며, 이를 기반으로 하는 번역은 정확하고 감정적 맥락을 반영하여 소통할 수 있는 장점이 있다[4]. 또한, Google Photos나 BeReal 같은 기존의 공유 앨범 플랫폼은 사진과 비디오 공유에 강점이 있지만 가족 간의 정서적 유대감을 강화하는 스토리텔링 기능은 부족하다. CODA와 농인 부모 간의 일상적 소통을 지원하는 시각적 소통 도구나 그림 생성 기능을 지원하지 않는다. 따라서, 수어 번역 기술과 공유 앨범 플랫폼을 결합한 새로운 플랫폼은 가족 간의 소통을 보다 깊이 있게 만들고, 정서적 유대감을 강화할 수 있다. 이 플랫폼은 수어, 텍스트, 음성 등의 다양한 입력 방식을 지원함으로써 가족 내 소통 문제를 해결하는 데 도움이 될 것이다.

본 논문에서는 CODA를 위한 AI 기반의 소통 플랫폼인 CODA 서비스 애플리케이션을 제안한다. 이 애플리케이션은 수어와 텍스트를 활용해 가족 간의 소통을 돕고, 일상을 그림으로 기록하는 기능을 제공함으로써 CODA 가족들을 정서적으로 연결한다. 논문의 구성은 다음과 같다. 2장에서는 기존의 수어 번역 및 공유 앨범 애플리케이션을 분석한다. 3장에서는 제안하는 애플리케이션의 설계와 주요 기능을 설명하며, 4장에서는 수어 번역 기술의 핵심인 Pose2Gloss, Gloss2Text, 이미지 생성 모델 구현에 대하여 기술한다. 5장에서는 애플리케이션 구현과 실험 결과를 설명하고, 마지막으로 6장에서 결론을 제시한다.

II. Preliminaries

1. Bench-marking

현재 수어를 인식하고 이를 텍스트로 변환하는 등의 서비스는 매우 제한적이다. 기존의 수어 번역 및 공유 앨범 애플리케이션 및 서비스를 분석한 결과는 표 1과 같다.

Table 1. Bench-marking of Sign to Language

Application Name	Advantage	Weakness
Google Sign Language Translator [5]	Converts sign language to text	Lacks support for other sign languages
	Accessible through multiple devices and platforms	
	AI Service	
HandTalk [6]	Specialized in Brazilian Sign Language (Libras)	Lacks support for other sign languages
	Widely used in public services and businesses in Brazil	Primarily text-based communication
	AI Service	
SignAll [7]	Supports American Sign Language (ASL)	Lacks support for other sign languages
	Provides feedback on user performance	Primarily text-based communication
	AI Service	

표 1에서 Google Sign Language Translator는 수어를 텍스트로 변환하는 기능을 제공하고, 다양한 디바이스와 플랫폼을 지원한다. 하지만 다른 시각적 커뮤니케이션 지원이 부족하다. HandTalk과 SignAll는 수어 인식 및 번역 기능을 제공하지만, 가족 간의 정서적 소통을 위한 시각적 도구나 그림 생성 기능이 부족하다. 특히 HandTalk은 Google Sign Language Translator로 수어를 텍스트로 변환한다. 하지만 텍스트 소통에만 국한되어 있어 CODA와 청인 가족이 함께 사용할 수 있는 시각적 커뮤니케이션 방법이 부족하다.

공유 앨범(다이어리) 기능에 대한 분석은 표 2와 같다. 기존의 앨범 앱들은 사진과 비디오를 쉽게 공유할 수 있는 장점을 가지고 있지만, 수어 사용자를 위한 스토리텔링 기능이나 가족 간의 감정적 연결을 제공하는 데에는 한계가 있다. BeReal과 Clubhouse와 같은 소셜 앱들은 실시간 상호작용에 중점을 두고 있지만, 가족 중심의 일상 기록이나 정서적 유대감을 강화하는 데 적합하지 않다. 따라서, 수어 번역과 공유 앨범을 결합한 플랫폼은 가족 간의 정서적 유대감을 강화하고, 다양한 소통 방식을 지원하여 기존 프로그램들이 해결하지 못한 문제들을 보완할 수 있다.

Table 2. Bench-marking of shared album

Application Name	Advantage	Weakness
Google Photos [8]	Easy media sharing with cloud backup	Primarily designed for media storage, lacks personal storytelling features
	Organizes photos automatically using AI	Lacks real-time interaction or emotional sharing tools
	Supports collaborative albums with multiple users	Limited to static media, lacks dynamic or interactive elements
BeReal [9]	Encourages spontaneous, real-time photo sharing	Limits posting to once a day, reducing flexibility for frequent updates
	Focuses on authentic moments by limiting daily posts	Lacks functionality for organized, long-term family album creation
	Engages users in a more personal and honest way	Primarily focuses on real-time, instant content rather than curated stories
Clubhouse [10]	Allows voice-based, real-time conversations	Primarily audio-based, lacks visual or media sharing capabilities
	Engages users with live, interactive discussions	Focuses on ephemeral, live interactions rather than permanent content
	Fosters a sense of community through live voice rooms	Limited to real-time discussions, lacks long-term diary or family album creation

2. Related Works

글로스는 수어 사용자가 단어의 조합 대신 의미적 최소 단위로 의사소통하는 방식으로, 단순히 완전한 문장 대신 핵심적인 의미 요소를 전달하는 것이 특징이다. 이러한 글로스 기반의 소통을 고려한 변환 기술이 도입되면 수어 사용자의 의사소통이 오류 없이 전달될 수 있다. 이 기술적 체계는 수어, 텍스트, 음성 등 다양한 입력 방식을 유연하게 변환할 수 있도록 도와줄 뿐만 아니라, 수어 번역 기술의 최대 난제 중의 하나인 데이터 부족 문제를 해결하는 데 기여할 수 있다. 특히, 수어는 상황에 따라 다양한 방식으로 표현될 수 있어 모든 상황에 적합한 데이터 수집이 어렵다. 그러나 글로스 변환 기술을 도입하면, 적은 데이터로도 효율적인 번역 시스템을 구축할 수 있으며, 다양한 의사소통 방식에 대응할 수 있는 환경을 제공할 수 있다.

Sign Language Recognition(SLR)은 수어 영상에서 글로스를 추출하기 위한 연구 분야이다. 컴퓨터 비전과 자연어 처리 기술의 발전과 함께 많은 연구가 이루어지고 있으며, 관련 연구들은 대체로 비디오 데이터에서 수어를 인식하고 이를 텍스트로 변환하는 방식에 중점을 두고 있으며, 이 과정에서 다양한 기술적 접근이 시도되고 있다.

첫째는 초기 연구들은 주로 정적 수어 이미지나 간단한 수어 제스처를 인식하는 데 초점을 맞추고 있다[11]. 휴대용 장치에서 수어를 실시간으로 인식하기 위해 Hidden Markov Model(HMM)을 활용한다. HMM은 시간에 따른 제스처의 연속성을 모델링하는데 효과적이지만 복잡한 수어 문장 인식에 한계가 있었다. 둘째는, Convolutional Neural Networks(CNN)와 Recurrent Neural Networks(RNN)의 발전은 SLR의 성능을 크게 향상시켰다. 특히 CNN은 이미지에서 특징을 추출하는 데 강점을 보였고, RNN은 시계열 데이터의 처리에 유리한 구조로 수어 문장 인식에 중요한 역할을 했다. 그리고 CNN을 사용해 손 모양을 인식하고, RNN을 사용해 시계열 데이터를 처리하는 모델을 제안하여 수어 인식 성능을 개선하였다[12]. 최근에는 Transformer 기반 모델이 SLR에 도입되면서 더욱 정확한 인식이 가능해졌다. Self-attention 메커니즘을 통해 긴 시퀀스의 수어 데이터를 효과적으로 처리하여 자연스러운 수어-텍스트 번역을 구현하였다[13]. 이러한 Transformer 기반 모델은 복잡한 수어 문법 구조를 처리하는 데 유리하다. 또한, 멀티모달 데이터를 활용한 연구도 활발히 진행되고 있다. 영상 데이터뿐만 아니라, 깊이 센서, 골격 데이터, 그리고 글로브형 센서 데이터를 함께 사용하는 방식은 수어 인식의 정확도를 높이는 데 기여하고 있다. 그리고 골격 정보를 활용한 수어 인식 연구를 통해, 손과 팔의 움직임을 보다 정확하게 추적함으로써 복잡한 수어 표현을 효과적으로 인식한다[14].

Sign Language Translation(SLT)은 수어를 자연어로 번역하거나 그 반대로 자연어를 수어로 번역하는 기술로 번역 결과의 자연스러움을 향상시키기 위한 자연어 처리기술을 함께 사용한다. 이에 Gloss-to-Text (Gloss2Text) Task는 수어 번역 분야에서 중요 작업으로 수어의 글로스 표현을 텍스트로 변환하는 과정이다. 글로스는 수어의 각 수어 동작을 설명하는 문자 표현으로, 일반적으로 단어 형태로 구성된다. 기존 연구에서는 AIHUB에 제공된 수어 데이터셋인 NIASL2021은 주로 재난 및 안전 정보 전달을 위한 수어에 초점을 맞추고 있어 데이터의 다양성에 제한이 있었다[15]. 따라서 응급 상황, 공항, 교통, 일상 4가지 영역의 샘플로 구축된 데이터인 GKSL(KETI-Emergency, Airport, Daily, NIA-2020)[16]를 채택하였다. KoBART 모델로 실험을 진행했고 BLUE 평가 지표로 성능을 측정해 준수한 성능을 나타내었다. 수어 해석에서 나타나는 다양성의 부재와 비언어적 요소인 얼굴 표정 및 상체 자세와 같은 비수지 신호의 중요성을 고려하지 않는 한계에도 불구하고, 중요한 진전을 보여주어 수어 연구가 더욱 발전할 것으로 예상된다.

III. Design of Recording CODA Service Application

1. CODA application architecture

CODA 서비스 애플리케이션은 개인 앨범, 공유 앨범, 이미지를 통한 소통 기능을 제공한다. 사용자인 농인과 CODA는 수어, 글로스, 텍스트, 음성의 네 가지 입력 방식 중 하나를 선택하여 자신의 일상을 기록한다. CODA 서비스 애플리케이션 구조는 그림 1과 같다.

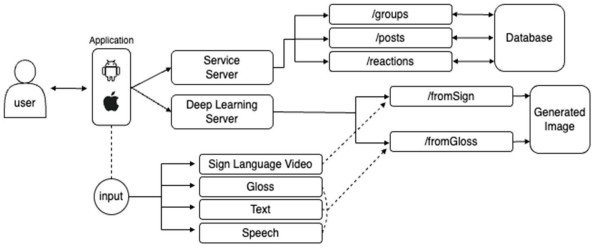

Fig. 1. CODA service application architecture

그림 1의 CODA 서비스 애플리케이션은 농인과 CODA가 함께 공유할 수 있는 감각인 시각을 통해 소통하도록 입력 내용을 4컷의 생성형 이미지로 변환하는 과정을 포함한다. CODA 서비스 애플리케이션 서버는 공유 앨범 서비스 기능을 제공하는 Service server와 입력 내용을 생성형 이미지로 변환하는 기능을 제공하는 Deep Learning Server로 구성된다.

2. Service Server

Service Server는 공유 앨범 서비스 기능을 위해 '/groups', '/posts', '/reactions' 엔드 포인트를 제공한다. '/groups' 엔드 포인트에는 공유 그룹 생성하기, 공유 그룹 참여하기, 공유 그룹 나가기, 공유 그룹 내 포스트 목록 조회하기 등의 세부 기능이 포함되며 이를 통해 사용자는 자신의 그룹을 관리한다. '/posts' 엔드 포인트에는 포스트 생성하기, 포스트 삭제하기, 포스트 정보 조회하기, 포스트 수정하기 등의 세부 기능이 포함되며 이를 통해 사용자는 각 공유그룹에 원하는 포스트를 공유할 수 있다. '/reactions' 엔드 포인트에는 리액션 보내기, 리액션 삭제하기, 리액션 조회하기 등의 세부 기능이 포함되며 이를 통해 사용자는 각 포스트에 대해 원하는 리액션을 보내며 포스트에 대한 감상을 함께 공유할 수 있다.

3. Deep Learning Server

Deep Learning Server는 사용자 입력을 기반으로 네컷의 이미지를 생성하여 반환한다. 이때 입력은 수어, 글로스, 텍스트, 음성의 4가지 형태이나 Deep Learning Server의 엔드 포인트는 '/fromSign',과 '/fromGloss' 2가지이다. 이는 글로스, 텍스트의 경우 동일한 과정을 거치고, 음성 입력 역시 front-end에서 STT(Speech To Text) 모듈을 통해 텍스트로 변환한 후 back-end로 전달되기 때문이다. 포스트 생성 과정에서 '/fromSign', '/fromGloss' 엔드 포인트가 호출되며, 이를 통해 front-end에서는 생성된 4컷의 이미지를 응답받는다.

IV. Method of Sign Language Translation

본 논문에서는 수어 영상을 입력받아 내용을 분석하고 이를 이미지로 변환하는 방법을 제안한다.

그림 2는 Image2Pose, Pose2Gloss, Gloss2Text, Image Generation 총 4개의 모듈로 구성된다. Image2Pose는 이미지에서 신체 키 포인트를 검출하는 과정이다. 이는 Google에서 개발한 mediapipe 라이브러리[18]를 이용해서 추출한다. Pose2Gloss는 연속적인 신체 키 포인트로부터 글로스를 추출하는 과정이다. 이는 그래프 합성곱 신경망 기반의 딥러닝 네트워크를 통해 구현한다. Gloss2Text는 앞 모듈에서 추출된 gloss를 text로 번역하는 과정이다. 해당 번역 Task는 한 문장 단위의 텍스트를 생성하는데 강점이 있는 한국어 버전의 T5(Text-To-Text Transfer Transformer) 모델인 KoT5를 통해 구현한다. Image Generation은 text를 image는 변환하기 위해 Open AI의 Dall-E 3 모델의 API를 사용하고, 프롬프트 튜닝을 통해 적절한 이미지를 생성한다.

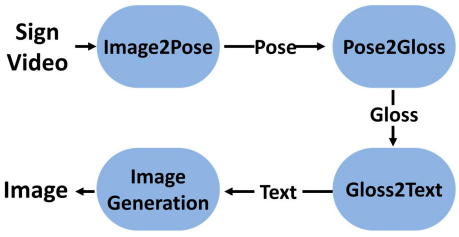

Fig. 2. Overall architecture of the sign language translation

1. Pose2Gloss

Pose2Gloss 모듈은 연속적인 Pose 키 포인트를 입력받아 어떤 의미의 Gloss인지 번역하는 기능을 한다. 본 논문에서는 연속적인 키 포인트의 시공간적인 정보를 고려하기 위해 GCN과 CNN을 이용한 모델과 수어 추가 기능을 위한 새로운 학습 방법을 제안한다. Pose2Gloss 모델을 도식화하면 그림 3과 같다.

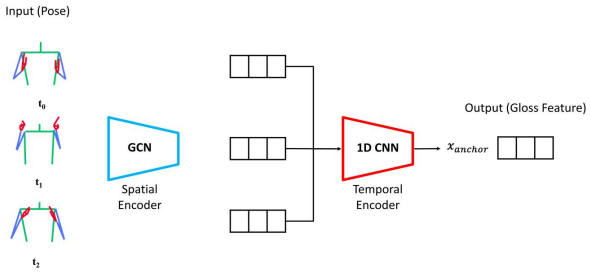

Fig. 3. Pose2Gloss model architecture

포인트의 공간적인 정보를 고려하기 위해 GCN을 이용하여 Pose 키포인트를 인코딩하여 특징을 추출한다[19]. 이어서 시간적인 정보를 고려하기 위해 시간 윈도우를 설정하여 일정 시간 동안의 pose 특징을 모은다. 이를 1D CNN 레이어 입력으로 하여 특징을 출력한다.

수어 추가 기능을 위해 Contrastive Learning을 통해 학습하고 Gloss를 추출하는 추론 방법을 제안한다. 그림 4 (a)는 Pose2Gloss 모델을 학습하기 위한 방법으로 Contrastive Learining을 이용해서 학습한다[20].

$$L(x_{anchor}, x_{positive}, x_{\neg ative}) = \| x_{anchor} - x_{positive} \|_2 + ReLU(Margin - \| x_{anchor} - x_{\neg ative} \|_2) \quad \text{(식 1)}$$

식 1은 Contrastive Learning을 위한 손실함수로 유클리디안 거리를 이용하여 Contrastive Loss를 정의한다. 이를 통해 Positive 특징은 서로 거리가 근접하도록 유도하고, negative 특징은 거리가 멀어지도록 유도하여 특징이 구별되도록 추출한다.

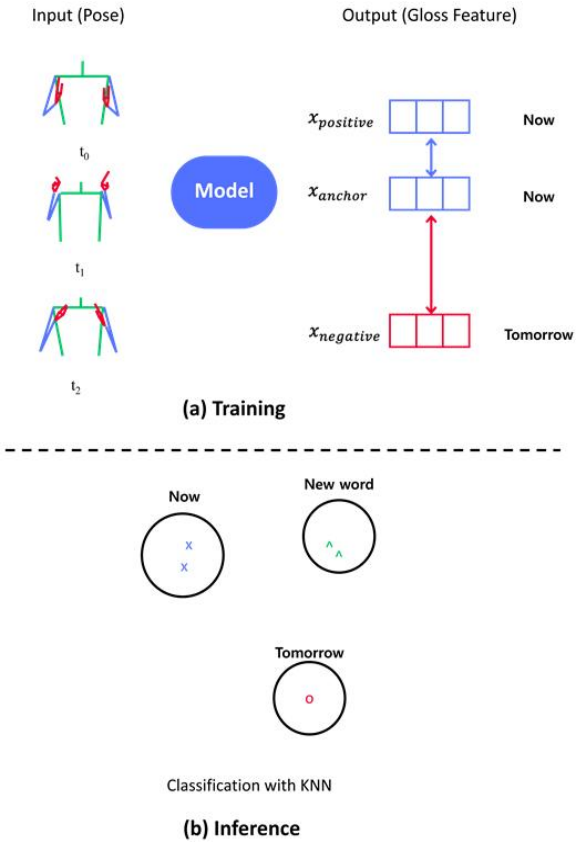

Fig. 4. (a) Training architecture (b) Inference architecture

그림 4 (b)는 추론 과정을 도식화한 것이다. 추론 과정은 수어 데이터셋의 특징 벡터와 입력 영상에서 추출한 특징 벡터를 비교하는 방식으로 구현되고, KNN을 이용하여 입력된 수어를 클러스터링하는 방식으로 gloss를 추출한다. 이를 구현하기 위해 모든 수어 영상 데이터에서 추출한 특징 벡터를 VectorDB로 저장한다. 이후 수어를 추가하는 경우 VectorDB에 새로 추가된 수어의 특징 벡터를 추가한다. 이를 통해 단순히 VectorDB에 저장하여 새로운 수어를 분류한다.

2. Gloss2Text

T5는 구글에서 개발한 대규모 사전 학습된 트랜스포머 기반 모델로, 텍스트 입력을 받아 다양한 자연어 처리 작업을 텍스트로 출력하는 구조이다[20, 21]. 다음과 같은 기준으로 모델을 선정한다.

- 번역 Task에 적합한 모델인가?
- 한국어로 사전학습되어 있는가?
- 모델 용량이 그리 크지 않고 학습 속도가 빠른가?

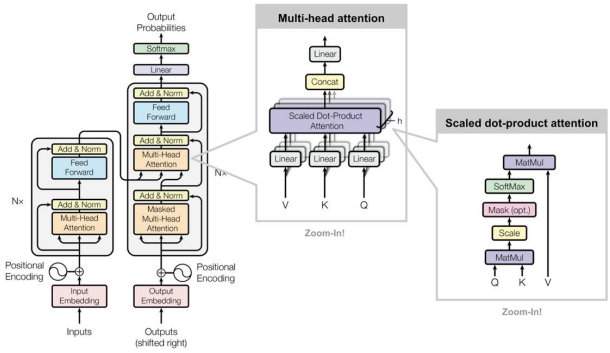

Fig. 5. T5 model architecture based on Transformer

이에 한국어 전용 데이터(나무위키, 위키피디아, 모두의 말뭉치 등)로 사전 학습된 KoT5를 해당 task에 Fine-tuning 하는 방식으로 학습한다. T5 모델은 그림 5와 같다. T5 모델은 농인 소통 방식에 별도의 학습을 수행하지 않았기 때문에 정확도가 높지 않다는 문제점이 있다. 따라서 본 논문에서는 KETI에서 제공하는 Gloss level Korean Sign Language 데이터(GKSL-dataset)를 활용해 번역 Task를 수행할 수 있도록 모델의 추가적인 학습을 수행한다. 이때, 사용자의 일상을 기록하는 서비스 특성에 맞게 의문문을 제외한 평서문만 포함해 크기가 8K인 데이터셋을 채택한다. 학습에 사용된 파라미터는 표 3와 같다.

Table 3. Simulation Parameter

Parameter	Value
Size of Model	77M
The Number of Dataset	8K
Epoch	87
Batch Size	16
Learning Rate	2e-5

Table 4. Comparison of performance between the convention Gloss2Text model and the fine-tuned model.

Model	BLEU	METOR
KoBART	33	0.27
KoT5_small	72	0.76

표 4는 기존 연구에 사용되었던 KoBART 모델과 본 논문에서 수행한 KoT5 모델의 성능을 BLEU(Bilingual Evaluation Understudy Score)와 METOR(Metric for Evaluation of Translation with Explicit ORdering)를 통하여 비교한 결과이다. KoT5가 KoBART 보다 가벼운 모델임에도 좋은 성능을 보인다.

3. Image Generation

이미지 생성 과정에서는 Open AI의 Dall-E 3 모델을 사용한다. Open AI의 API를 프롬프트 튜닝을 거쳐 목적

에 맞는 이미지가 생성한다. 프롬프트에 반영된 사항은 그림 6과 같다.

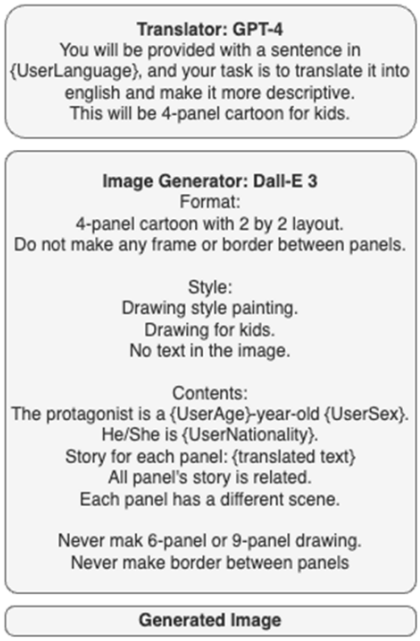

Fig. 6. Prompt Sample

그림 6의 프롬프트에 따라 4컷의 연속적인 스토리가 담긴 이미지를 생성하면 그림 7과 같다. 이때 동일한 그림체로 이미지가 생성될 수 있도록 2×2 배열의 이미지를 생성하도록 한다. 아동이 포함된 가족이 주된 타겟이기에 그림 스타일로 지정하고, 아이를 위한 이미지임을 명시한다. 텍스트를 잘 반영하지 못하는 생성형 이미지의 한계점을 고려하여 그림에 텍스트를 포함하지 않도록 한다. 일상을 기록하기 위한 용도이기에 사용자의 연령, 성별, 신체 특징 등을 프롬프트에 포함한다. 반복적인 테스트를 통해 프롬프트가 잘 반영되지 않는 부분은 동일한 내용을 두 번 중복하여 작성한다.

Fig. 7. 4-cut image generated by prompt

V. Implementation of CODA Service Application

CODA 서비스 애플리케이션 구현을 위한 소프트웨어 및 프레임워크는 표 5와 같다.

Table 5. Development Language & Framework

Usage	Frame Work
Front End	Flutter
Back End	FastAPI
Data Base	MySQL
API	Open AI, Mediapipe
Deep-Learning	Python, PyTorch

1. Main page

CODA 서비스 애플리케이션의 메인 화면은 그림 8과 같다.

Fig. 8. Main page

그림 8에서는 개인 앨범과 공유 앨범을 리스트 형태로 확인할 수 있다. 화면 하단의 버튼을 누르면 새로운 공유 앨범을 생성하거나 이미 생성된 공유 앨범에 초대 코드를 통해 입장할 수 있다.

2. Album

그림 8에서 각 앨범을 클릭하면 그림 9의 화면이 나타난다.

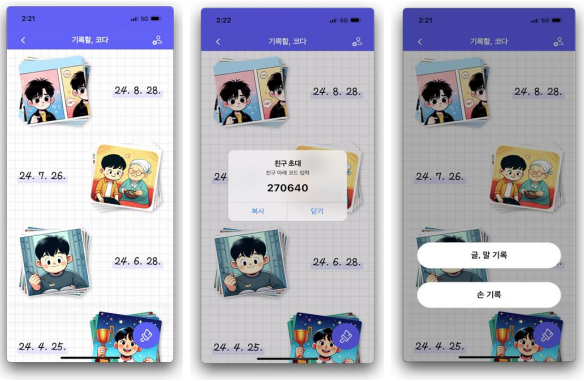

Fig. 9. 4-cut image generated by prompt

그림 9에서는 공유 앨범의 제목이 나타나고, 구성원은 상단 오른쪽 버튼을 통해 초대 코드를 생성 및 공유할 수 있다. 발행된 코드를 다른 사용자가 그림 9의 메인 화면에 입력하면 해당 앨범에 초대된다. 화면에 생성된 4컷 이미지는 날짜를 기준으로 내림차순으로 표시한다. 화면 오른쪽 하단의 버튼을 클릭하면 2가지(수어 영상과 글로스) 형태의 입력으로 일상을 기록할 수 있다.

3. Input story

그림 10 (a)에서는 일상을 수어 영상으로 입력한다.

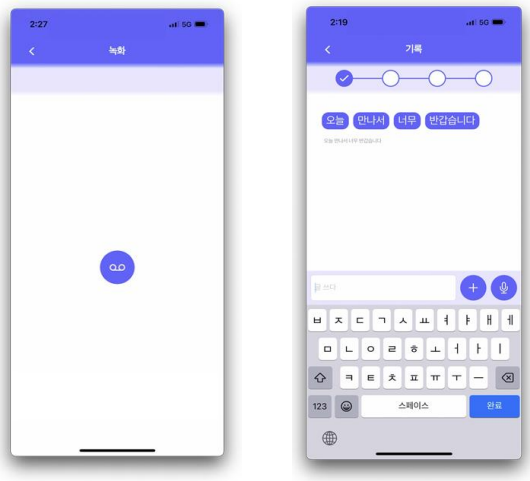

Fig. 10. 4-cut image generated by prompt

화면 중앙부 재생 버튼을 누르면 3초의 타이머 후에 녹화를 진행한다. 녹화 전에는 사용자의 상체 부분의 랜드마크 일부를 시각적으로 확인이 가능하도록 하여 위치를 조정할 수 있다. 8초간 녹화 후에는 다음 녹화를 시작하거나 입력을 완료할 수 있는 버튼 2개가 수평으로 나란히 놓인다. 입력은 총 4번 가능하고, 중간에 입력을 종료할 수 있다. 그림 10(b)에서는 글로스, 문장, 그리고 음성 인식으로 일상을 입력한다. 화면 중앙은 메신저 속 채팅 형태로 구성한다. 입력 창에서는 글로스, 문장, 그리고 입력 창 오른쪽 음성 마이크 버튼을 통해 음성 입력을 할 수 있다. 입력은 총 4번이 가능하고 중간에 입력을 마칠 수 있다. 그리고 입력 일자를 선택할 수 있고, 앨범에 접근하여 사진을 첨부할 수 있다. 모든 입력을 완료 후에 화면 아래 만들기 버튼을 클릭하여 입력 정보를 Back-end 서버로 전달한다. 요청과 응답을 완료하고, feed 입력하면 그림 11과 같은 화면을 생성한다.

Fig. 11. Feed input

VI. Conclusions

본 논문에서는 생성형 AI 기반의 CODA 서비스 애플리케이션을 설계하고 구현하였다. 이 애플리케이션은 CODA 가족을 위한 소통 플랫폼으로 수어 번역, Gloss 번역, 이미지 생성 기술을 적용하여 구현한다. 온라인 공유 앨범을 통해 가족 내 공감대를 형성하고, 그 과정에서 수어, 글로스, 텍스트, 음성 등의 4가지 입력이 가능하도록 구현함으로써 입력의 편리성을 향상하였다. 또한, 수어 번역 기능에 사용된 각 인공지능 모듈은 개별 기능으로 확장 가능하다. Pose To Gloss 모듈은 수어 교육 기능으로 확장될 수 있으며 Gloss To Text 모듈은 문장 위주의 텍스트에 어려움을 느끼는 농인들을 위한 채팅 기능으로 확장될 수 있다. 그리고 Pose To Gloss 모듈의 경우, 모델 학습 없이 DB 구축만으로 동작하기 때문에 신조어 등 새로운 수어 단어에 능동적으로 대응할 수 있으며 더 나아가 다양한 언어에 추가 학습 없이 적용할 수 있다. 이러한 확장성은 CODA 서비스 애플리케이션의 성능 향상에 도움이 될 것으로 기대한다.

ACKNOWLEDGEMENT

This work was supported by the SK Telecom's FLY AI Challenger program, conducted in collaboration with the Ministry of Employment and Labor and the Korean Skills Quality Authority as part of the 2024 K-Digital Training.

REFERENCES

[1] Jenny L. Singleton, Matthew D. Tittle, "Deaf parents and their hearing children," Journal of Deaf Studies and Deaf Education, Vol. 5, No. 3, pp.221-236 July 2000 https://doi.org/10.1093/deafed/5.3.221

[2] https://www.deafkorea.com/main/

[3] Claudia M. Pagliaro, "Technology use by deaf individuals," Journal of Deaf Studies and Deaf Education," Vol. 19, No. 3, pp.112-120 Mar. 2014 https://doi.org/10.1093/deafed/enu005

[4] Sutton-Spence, R., & Woll, B., "The Linguistics of British Sign Language: An Introduction." Cambridge University Press. 1999 https://doi.org/10.1017/CBO9781139167048

[5] https://play.google.com/store/apps/details?id=com.siyaroll.sltranslator

[6] HandTalk : https://www.handtalk.me

[7] SignAll : https://www.signall.us

[8] Google Photos : https://photos.google.com

[9] BeReal : https://bere.al

[10] Clubhouse : https://www.clubhouse.com

[11] Starner, Thad, Joshua Weaver, and Alex Pentland. "A wearable computer based american sign language recognizer," Digest of Papers. First International Symposium on Wearable Computers. IEEE Xplore, 1997. https://doi.org/10.1109/ISWC.1997.629929

[12] Pigou, Lionel, et al. "Sign language recognition using convolutional neural networks," Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13. Springer International Publishing, pp. 572-578 Jan. 2015.

[13] Necati Cihan Camg", Oscar Kollerq, Simon Hadfield and Richard Bowden, "Sign language transformers: Joint end-to-end sign language recognition and translation," Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[14] Koller, Manuel. "robustlmm: an R package for robust estimation of linear mixed-effects models," Journal of statistical software, Vol. 75 No. 6, pp. 1-24, Dec. 2016 https://doi.org/10.18637/jss.v075.i06

[15] J. K. Oh, Y. C. Kim, M, G, Jeon. "Enhancing Korean Sign Language Translation with Synthetic Data Generation via Large Language Model," Proceedings of the Korea Software Congress 2024

[16] "Gloss level Korean Sign Language dataset," https://github.com/AIRC-KETI/GKSL-dataset , KETI, 2024

[17] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg and Matthias Grundmann, "Mediapipe: A framework for perceiving and processing reality." Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR) 2019.

[18] Thomas N. Kipf, Max Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016. https://doi.org/10.48550/arXiv.1609.02907

[19] Zhirong Wu, Yuanjun Xiong, Stella Yu, Dahua Lin, "Unsupervised feature learning via non-parametric instance discrimination." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. https://doi.org/10.48550/arXiv.1805.01978

[20] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J. Liu, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," Journal of Machine Learning Research, Vol. 21, No. 1, pp. 5485 - 5551 Jan. 2020

[21] Paust, "pko-t5-small. Hugging Face," https://huggingface.co/paust/pko-t5-small 2021.

Authors

Jeongmin Ahn will receive the B.S. degree in Statistics and IT Engineering(double major) from Sookmyung Women's University, Korea, in 2025. Her research interests include natural language processing, large language model, and data analysis.

Hyewon Ryu will receive the B.S. deree in Industrial and Management Engineering from Korea University, Korea, in 2025. Her research interests include natural language processing, graph neural network, recommendation system, and data analysis.

Dusan Baek received a B.S. degree in Material Science and Engineering and Computer Science and Engineering (double major) from the Ulsan National Institute of Science and Technology (UNIST), Korea, in 2024. His research interests focus on Human-Computer Interaction, User Experience, and Large Language Models.

Gyeongho Cho received the B.S. and M.S. degrees in Mechanical Engineering from Pusan National University, Korea, in 2022 and 2024. His research interests include perception and control for the autonomous driving.

Seongbin Choi will receive the B.S. degree in Geography with a Double Major in Computer Engineering from Kyunghee University, Korea, in 2025. His research interests include Spatial Data Analysis, Natural Language Processing, Geographic Information Systems, Geospatial Programming, and Machine Learning Applications.

Ho-Young Kwak received the B.S., M.S., and Ph.D. degrees in Computer Science from Hong-Ik University, Korea, in 1983, 1985, and 1990, respectively. Dr. Kwak joined the Department of Computer Engineering at Jeju National University, Jeju, Korea, in 1990. He is currently a Professor in the Department of Computer Engineering, Jeju National University. He is interested in IT-Medical convergence, Companion Animal Healthcare systems, and Software systems.

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.