컴퓨터교육학회 논문지 2024년 제27권 제8호
https://doi.org/10.32431/kace.2024.27.8.009

YOLOv8을 이용한 한글 지문자 인식*

Korean Fingerspelling Recognition Using YOLOv8

김진영† · 강의성†† · 장문수†††
Jinyoung Kim† · Euisung Kang†† · Moonsoo Chang†††

요약

청각 장애인은 음성을 사용할 수 없으므로 주로 수어를 통하여 의사소통을 한다. 수어의 한 종류인 한글 지문자는 손과 손가락의 모양으로 한글의 자음과 모음을 나타낸다. 청각 장애인들이 지문자를 익히면 의사소통을 쉽게 할 수 있지만 배우기가 쉽지 않다. 본 논문에서는 인공지능 기술을 이용하여 수어 교육에 활용할 수 있는 지문자를 인식하는 알고리즘을 제안하고자 한다. 이를 위해 본 논문에서는 이미지 인식에 많이 활용되는 YOLOv8을 활용한 한글 지문자 인식 알고리즘을 제안한다. 또한 YOLO 학습에 사용되는 다량의 학습 데이터를 자동으로 구축하기 위하여 미디어파이프(MediaPipe)의 Landmark 기능을 활용한 자동 어노테이션 방법을 제안한다. 제안하는 방법의 성능을 기존 연구의 SVM(support vector machine) 알고리즘과 비교하여 우수성을 검증하였다. 특히, 오인식률이 높은 한글 지문자 'ㅋ','ㅊ','ㅋ','ㅋ'에 대해서도 높은 인식률을 보임을 확인하였다.

주제어 한글 지문자, YOLOv8, MediaPipe, SVM

ABSTRACT

Deaf people communicate with sign language using their hands. Korean fingerspelling, a type of sign language, represents Korean consonants and vowels through hand and finger shapes. Although fingerspelling is an effective communication method, it can be challenging to learn. This paper proposes an algorithm for recognizing fingerspelling to support sign language education using artificial intelligence technology. Specifically, we present a Korean fingerspelling recognition algorithm based on YOLOv8, which is widely used in image recognition. Additionally, we introduce an automatic annotation method using MediaPipe's Landmark feature to efficiently create the large amount of training data required for YOLO training. The performance of the proposed method was validated by comparing it with the Support Vector Machine (SVM) algorithm from previous research, demonstrating its superiority. Notably, our method achieved high recognition accuracy for Korean fingerspelling characters 'ㅋ,' 'ㅊ,' 'ㅋ,' and 'ㅋ,' which typically have high misrecognition rates.

Keywords Korean Fingerspelling, YOLOv8, MediaPipe, SVM

†정회원 국립순천대학교 컴퓨터교육과 시간강사
††정회원 국립순천대학교 컴퓨터교육과 교수
†††정회원 서경대학교 소프트웨어학과 부교수(교신저자)

논문투고 2024년 11월 07일
심사완료 2024년 11월 19일
게재확정 2024년 11월 20일
발행일자 2024년 11월 27일

* 본 논문은 순천대학교 교연비 사업에 의하여 연구되었음.

1. 서론

청각 장애인은 음성으로 표현하는 의사소통에 어려움이 있어 수어, 구화, 필담 등을 이용하여 의사소통을 한다. 2020년 국립국어원에서 실시한 한국수어 활용 조사에 의하면 청각 장애인들의 주된 의사소통 방법은 수어가 가장 높은 것으로 조사되었다[1].

수어는 청각 장애인들이 자신의 생각과 감정을 표현하고 타인과 소통할 수 있는 중요한 수단으로 음성이 아닌 동작을 통해 표현되는데, 손의 모양, 손바닥의 방향, 손의 위치, 손의 움직임, 얼굴 표정 등으로 생각을 표현하는 의사소통 방법이다[2]. 수어의 한 종류인 지문자는 손과 손가락의 모양으로 한글의 자음과 모음을 나타낸 비음성 언어이다. 지문자는 사람 이름, 장소 이름, 기관명 등의 고유 명사를 표현할 때 사용하며, 전문적인 단어나 기술 용어를 전달할 때 유용하다[3].

Karen과 Jennifer[4]의 연구에서 지문자 인식 능력과 영어 단어 읽기 능력 사이에 강한 상관관계가 있음을 밝혔다. 특히, 지문자 능력은 읽기 이해력 및 어휘력과 유의미한 상관관계가 있으며, 이는 지문자가 단순히 수어의 보조 도구가 아닌, 청각 장애인들의 언어 발달의 핵심 요소임을 시사한다.

지문자 인식에 관한 연구로는 손가락 개수와 손가락 방향을 파악하여 실시간으로 한글 지화 번역을 위한 CNN 모델을 활용한 연구[5], 틸트 센서와 플렉스 센서를 이용한 한글 지화 인식용 데이터 글러브 시스템 연구[6], 아두이노(Arduino), Flex Sensor, AHRS Sensor를 이용하여 손의 움직임을 파악하는 장갑 형태의 수어 번역기 구현에 관한 연구[7], 한글 지문자 자음을 대상으로 화상 카메라 영상에서 피부 영역을 추출하고 윤곽선 추적 알고리즘으로 손 후보 영역을 추출하는 지화 인식 연구[8], 적외선 카메라 입력 영상을 처리 및 분석하여 정보기기 제어를 위한 지화 인식 인터페이스 시스템을 개발하는 연구[9], YOLOv7 알고리즘을 이용하여 지문자와 특정 단어를 대상으로 한 실시간 수어 번역 AI 프로그램 구현에 관한 연구[10], CNN과 LSTM을 결합한 모델을 이용하여 수어 동작을 판별하는 실시간 플랫폼 개발에 관한 연구[11]등이 있다.

그런데 지문자 인식에 관한 선행 연구 중 상당수는 지문자 교육에 활용하기에 한계가 있었다. 총 31개 한글 지문자가 아닌 일부 지문자만을 대상으로 하거나 지문자에 대한 인식률이 낮아서 실제 교육에 적용하기 어려운 경우도 있었다. 그리고 일부 연구에서는 데이터 글러브 또는 적외선 카메라 등과 같이 별도의 장비가 필요하여 정해진 장소에서 활용하거나 휴대하여야 하는 제한이 있었다.

지문자 인식 결과를 지문자 교육에 활용하기 위해서는 모든 한글 지문자에 대해 인식이 가능하고, 신체에 부착하거나 착용할 필요가 없이 저렴하고 쉽게 접할 수 있는 웹카메라를 이용한 지문자 인식 기술이 필요하다.

본 논문에서는 객체 검출에 널리 사용되는 YOLO(you only look once)를 이용하여 한글 지문자를 인식하는 인공지능 모델을 구현하여 지문자 교육에 활용가능한 지문자 인식 알고리즘을 제안한다. 또한 제안하는 지문자 인식 모델의 성능을 검증하기 위하여 SVM(support vector machine) 알고리즘을 이용한 기존 연구의 모델과 성능을 비교한다.

본 논문의 구성은 다음과 같다. 2장에서는 지문자 인식에 사용되는 인공지능 알고리즘에 대해 소개한다. 3장에서는 한글 지문자 인식을 위한 인공지능 모델을 제안하고, 학습에 필요한 데이터셋 구축 알고리즘을 설명한다. 4장에서는 제안하는 모델과 비교 모델의 실험 결과를 나타내고 성능을 비교한다.

2. 관련 연구

2.1 지문자 인식을 위한 알고리즘

2.1.1 YOLOv8 알고리즘

YOLO는 컴퓨터비전 분야에서 실시간으로 객체를 검출하는 도구로 많이 사용하는 딥러닝 기반 모듈이다. 2016년에 첫 버전이 발표된 이후 현재까지 꾸준히 새로운 버전이 나오고 있다. 2023년 YOLOv8 버전이 안정적으로 서비스되고 있으며, 이후 계속 새로운 버전이 공개되고 있다[12]. 최근의 YOLO 버전은 초기 버전에 비하여 검출의 성능이 좋아지면서도 실시간 객체 검출의 목적에 맞춰 모델의 크기를 줄여서 속도도 같이 향상되고 있다. 안정된 버전으로 최신 버전에 속하는 YOLOv8은 이전 버전과 달리 앵커 프리 모델로 앵커 박스를 사용하는 이전 모델에 비하여 객체 예측의 수를 줄임으로써 속도를 개선하고 있다. 또한 첫 단계의 컨벌루션 커널을 1x1에서 3x3으로 변경하여 전체 파라미터 수를 줄임으로써 모델의 크기를 줄였다[13, 14].

본 논문에서는 YOLOv8 중에서 YOLOv8n 모델을 활용하여 한글 지문자 인식 알고리즘을 구현한다.

2.1.2 SVM 알고리즘

SVM은 Vladimir Vapnik과 벨 연구소 동료들에 의해 개발되었으며, 클래스 간 마진(margin)이 가장 높은 n차원 분류 공간에서 최적의 초평면을 찾는 방법이다[15]. SVM은 지도 학습에 기반한 인공지능 모델 중의 한 가지로 선형 분류와 비선형 분류 작업에 모두 사용될 수 있고, 작은 데이터 집합에도 유용하게 적용될 수 있다. 또한, SVM은 일반적으로 이진 분류 문제를 해결에 유용하게 활용될 수 있으며, 이진 분류 문제가 아닌 경우에도 여러 개의 이진 분류 문제로 나누어 각각을 해결하고 그 결과를 종합하여 결정하는 방식으로 사용할 수 있다[16].

SVM의 장점은 고차원 공간에서 효과적이고, 차원 수가 표본 수보다 많은 경우에도 효과적이다. 또한, 결정 함수에서 훈련 포인트의 부분 집합(support vectors)을 사용하므로 메모리 효율이 좋은 장점이 있다[17].

2.2 MediaPipe Hands를 이용한 지문자 인식

미디어파이프(MediaPipe)는 구글의 오픈 소스 라이브러리로서 손(hand), 우리 몸의 자세(pose), 얼굴(face) 등에 특징점을 검출하는데 뛰어난 성능을 보인다[18]. 손, 자세, 얼굴 등에 존재하는 특징점을 Landmark라고 하는데, Fig. 1은 미디어파이프 Hands에 대한 21 개의 Landmark를 보여주고 있다.

지문자는 특정한 손가락을 구부려지거나 펴진 모양으로 이용해서 한글 자음과 모음을 표현하기 때문에 위와 같은 Landmark 정보는 지문자를 인식하는 데 매우 유용하게 사용될 수 있다.

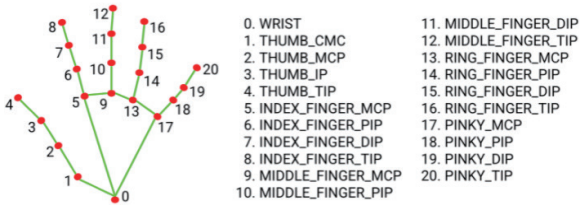


Figure 1. MediaPipe Hand Landmarks[18]

기존 연구[3]에서는 손 이미지로부터 추출된 Landmark를 학습 데이터로 하여 SVM을 이용하여 한글 지문자를 인식하는 학습 모델을 생성하였다.

3. 한글 지문자 인식 알고리즘

3.1 YOLOv8을 이용한 한글 지문자 인식 모델

SVM과 Landmark를 이용한 기존의 한글 지문자 인식 방법[3]은 31 개 지문자 중에서 대부분은 인식이 잘 되지만 손가락 끝이 카메라 정면을 향하는 특정 지문자 'ㅓ','ㅔ','ㅕ','ㅖ' 에 대한 인식률이 낮은 문제가 있었다. 이를 해결하기 위해 정면과 측면에 두 대의 카메라를 이용하였고 전체 지문자에 대한 인식률이 고르게 향상되었다.

그러나 이 방법은 정면과 측면 카메라의 교차점에서 손이 벗어나지 않아야 하기 때문에 손의 위치가 제한되고, 이런 점 때문에 실제 상황에서 활용하는데 제약이 따를 수 밖에 없다.

본 논문에서 제안하는 방법은 손의 위치를 비교적 자유롭게 유지하기 위하여 한 대의 카메라를 이용한다.

기존 방법에서 지문자 인식에 사용한 Landmark는 손 모양에서 21 개의 좌표값밖에 없기 때문에 정면 카메라의 영상에서 중복되는 좌표값을 가진 지문자는 인식에 불리할 수 밖에 없다.

본 논문에서는 정보가 부족한 Landmark 대신에 손의 이미지 전체 정보를 활용하기 위하여 YOLO를 사용하여 이미지를 인식하는 방식으로 지문자를 인식하고자 한다.

본 논문에서는 YOLOv8n을 사용하여 지문자를 인식한다. YOLOv8은 다양한 YOLO 버전 중에서 최신 모델에 속하며 YOLOv8n은 YOLOv8 중에서는 인식률이 좋으면서도 가장 계산량이 적어 속도가 빠르다는 장점이 있다.

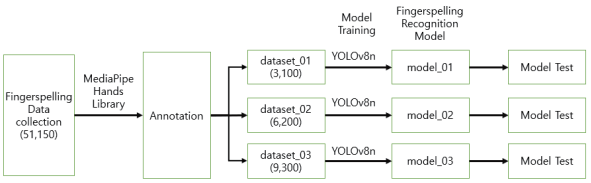


Figure 2. YOLOv8 based Fingerspelling Recognition model development process

Fig. 2는 YOLOv8n을 이용하여 지문자 인식 모델을 만드는 과정을 보여주며, 생성된 모델을 통해 지문자를 인식하고 모델의 성능을 분석하는 과정은 다음과 같다.

① 데이터셋 구축
- 어린이 2 명과 성인 9 명을 대상으로 수집한 총 51,150 개 지문자 이미지 준비

② 어노테이션 진행
- YOLO 모델 훈련을 위한 어노테이션 작업을 위해 미디어파이프를 이용하여 자동화

③ 학습용 데이터셋 준비
- 이미지 개수를 달리하는 3가지 데이터셋 준비

④ 인공지능 학습 및 모델 생성
- YOLOv8n을 이용한 인공지능 학습
- 3 개 지문자 인식 모델 생성

⑤ 모델 성능 분석
- 각 모델별 지문자 인식 테스트

제안하는 지문자 인식 모델을 위한 YOLOv8n의 하이퍼파라미터는 Table 1과 같이 설정한다.

Table 1. Training parameters

Hyperparameters	value
model	yolov8n
batch	-1
epoch	100, 300, 400, 500, 1000
imgsz	640

batch 파라미터의 '-1'은 자동 batch 크기를 의미하며 구동하는 컴퓨터의 메모리를 효율적으로 사용할 수 있도록 설정된다. imgsz는'640'으로 설정하였으며 이는 640×640 픽셀 크기의 정사각형 이미지를 의미한다. epoch 파라미터는 최적 모델을 찾기 위해 YOLO의 일반적인 권장사항인 300 외에도 100, 400, 500, 1000을 설정한다.

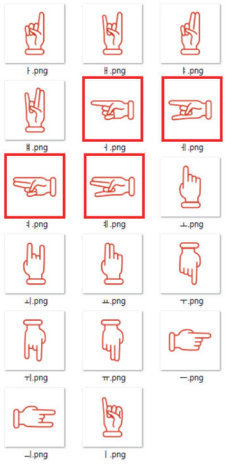
3.2 지문자 데이터셋 구성

한글 지문자는 Table 2와 같이 자음 14 개, 모음 17 개, 합계 31 개가 있다. 이 중에서 굵은 테두리의 4 개 문자

(‘ㅓ’,‘ㅖ’,‘ㅕ’,‘ㅞ’)는 손 모양의 방향 때문에 기존 연구에서 인식률이 떨어지는 문자들이다.

본 연구에서는 초등학생 2명, 성인 여성 5 명, 성인 남성 4 명을 대상으로 카메라 정면에서 손의 각도를 달리하며 30 fps(frame per second) 속도로 10 초 분량으로 촬영하였다. 총 11 명을 대상으로 촬영한 각 지문자 영상에서 150 개씩 이미지를 추출하여 총 51,150 개의 이미지를 생성하였다. 본 논문에서는 이 이미지들을 학습, 검증 및 테스트를 위한 데이터셋으로 사용한다.

Table 2. Korean Fingerspelling

consonant	vowel
	

또한, 데이터셋 크기에 따른 한글 지문자 인식 성능의 차이를 확인하기 위해서 Table 3과 같이 각 지문자별로 100 개에서 300 개까지 100 개 단위로 구성된 총 3 가지의 무작위 데이터셋을 구성한다. 예를 들어, dataset_01은 각 지문자 이미지 100 개씩, 총 3,100 개의 이미지로 구성되며, 학습, 검증, 테스트 데이터는 7:2:1의 비율로 설정하여 학습용 2,170 개, 검증용 620 개, 테스트용 310 개 이미지로 나누어 사용된다.

Table 3. Hands Datasets

Type	train (70%)	validation (20%)	test (10%)	total (100%)
dataset_01	2,170	620	310	3,100
dataset_02	4,340	1,240	620	6,200
dataset_03	6,510	1,860	930	9,300

3.3 미디어파이프를 이용한 손 영역 검출

YOLOv8은 객체 탐지, 인스턴스 분할, 포즈 추정, 분류, 다중 객체 추적과 같은 컴퓨터 비전 작업을 지원하는 여러 가지 데이터셋을 지원하고 있다[19]. 그러나 지문자 인식을 위한 데이터셋으로는 Roboflow에서 구축한 미국 수어(American Sign Language(ASL)) 데이터셋만 제공하고 있으며[20], 한글 지문자 데이터셋은 제공하지 않는다. 따라서 YOLO를 이용한 한글 지문자 인식에서는 자동으로 손모양 객체를 추적할 수 없기 때문에 추가적으로 어노테이션 작업이 필요하다.

데이터 어노테이션은 원시 데이터에 의미 있는 정보를 추가하는 작업으로 인공지능 모델이 데이터를 이해하고 학습을 할 수 있도록 각 데이터에 태그를 지정하는 과정이다. YOLOv8n을 이용해서 지문자 인식 모델을 생성하는 본 논문의 알고리즘에서도 이 과정은 매우 중요하다.

일반적으로 이미지 어노테이션 작업은 학습 데이터의 품질 보장을 위해 LabelImg[21]와 같은 도구를 사용하여 수동 어노테이션을 한다. LabelImg는 직관적인 그래픽 인터페이스를 제공하여 수동 어노테이션 작업을 용이하게 하고, YOLO 형식의 어노테이션 파일 생성을 지원한다.

그러나 LabelImg를 이용한 수동 어노테이션 방식은 정확도가 높은 반면, 학습 데이터가 많을 경우 상당한 시간과 노력이 요구된다. 본 논문에서 사용한 51,150개의 지문자 이미지에 대해서 수동으로 어노테이션을 하기 위해서는 많은 시간이 소요되며, 향후 지문자 인식 성능을 개선하기 위해서 새로운 학습 데이터를 수집할 때마다 어노테이션을 위해서 많은 시간을 할애해야 하는 어려움이 있다.

이러한 한계를 극복하기 위해 본 논문에서는 미디어파이프 Landmark를 이용하여 어노테이션 과정을 자동화한다. 손 이미지의 21 개 Landmark는 손 전체에 분포하고 있기 때문에 Landmark를 기준으로 전체 이미지 안에서 손 영역을 지정하여 어노테이션으로 활용할 수 있다.

제안하는 Landmark를 이용한 자동 어노테이션 과정은 다음과 같다.

① 원본 이미지의 너비와 높이를 0~1로 정규화
- 픽셀의 좌표를 원본 이미지의 너비와 높이로 나누어서 0~1 사이의 값을 갖도록 정규화

② Landmark 검출
- 미디어파이프 Hands를 이용하여 이미지 내에 존재하는 21 개의 손 Landmark 추출

③ 바운딩 박스 설정
- 추출된 21 개의 Landmark 좌표 중 x 좌표와 y 좌표에 대한 최소값과 최대값을 이용하여 바운딩 박스의 범위 결정
- x 좌표의 최소값과 최대값 (min_x, max_x)
- y 좌표의 최소값과 최대값 (min_y, max_y)

③ 바운딩 박스 중심점 계산
- 바운딩 박스의 중심점 x_center, y_center을 정규화된 좌표를 이용하여 계산

③ 바운딩 박스 크기 계산
- 바운딩 박스의 너비(width)와 높이(height)를 정규화된 좌표를 사용하여 계산

④ 어노테이션 파일 생성
- class ID, x_center, y_center, width, height 정보를 포함하는 어노테이션 파일 생성

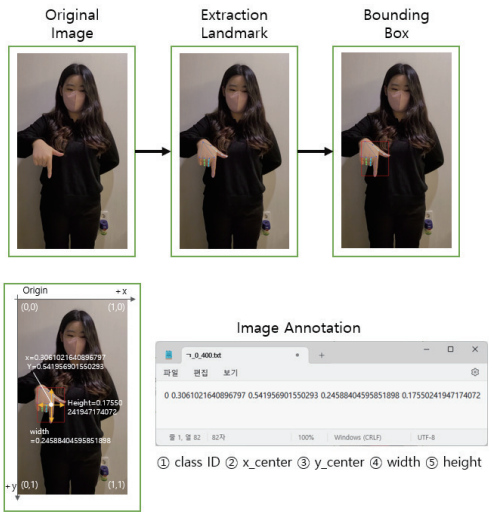


Figure 3. Training dataset construction process using MediaPipe Hands Library

Fig. 3은 앞에서 설명한 어노테이션 과정을 그림으로 나타내고 있다. 제안하는 자동 어노테이션 방법은 어노테이션 시간을 단축할 수 있을 뿐만 아니라 단순하고 반복적인 작업으로 인하여 발생할 수 있는 인적 오류의 가능성도 줄일 수 있다.

3.4 성능 비교를 위한 기존 알고리즘의 설정

제안하는 알고리즘과 성능을 비교하기 위하여 Landmark를 사용하는 기존의 방법을 본 논문의 학습 데이터에 맞춰서 재설정한다. Fig. 4는 MediaPipe Hands에서 제공하는 Landmark 정보와 SVM을 이용한 기존 학습 모델을 도식화하고 있다.

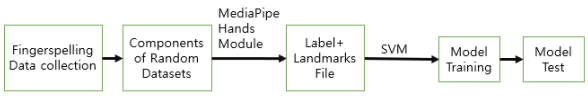


Figure 4. SVM based Fingerspelling Recognition model

객체 분류 알고리즘을 이용한 지문자 인식 모델을 만들기 위해 본 논문의 YOLOv8n 모델에서 사용하는 동일한 데이터셋을 이용하였다. 각 지문자 이미지 파일에 대한 정보는 지문자에 대한 Label, 손 영역에 대한 Bounding Box, 21 개의 정규화된 Landmarks 등을 Fig. 5와 같은 csv파일을 작성하여 SVM 모델 학습에 이용하였다.

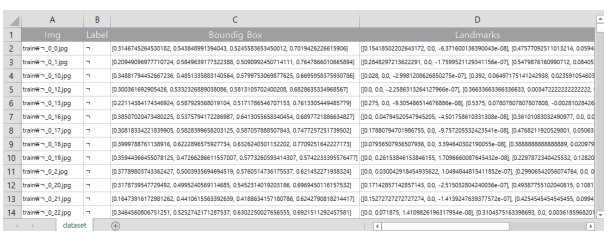


Figure 5. Label, Bounding Box, and Landmarks Data File

4. 실험 결과 및 평가

4.1 개발 환경

제안하는 알고리즘을 구현하기 위한 컴퓨팅 환경은 프로세서 i7-8086K 4.00GHz, RAM 64GB, 그래픽카드 NVIDIA GeForce RTX 2080(11GB RAM)으로 구성된 데스크톱 하드웨어에, OS는 Windows 10 x64를 사용하였다.

개발 소프트웨어는 Python 3.10.x 버전에서 cuda 11.2, scikit-learn 라이브러리 1.4.2, ultralytics 8.1, MediaPipe 0.10.11 버전을 사용하였다.

4.2 지문자 인식 모델의 성능 분석

제안하는 지문자 분류 모델의 성능을 분석하고 최적 모델을 도출하기 위하여 YOLO 모델을 평가하는 성능 지표인 mAP(mean Average Precision)를 사용한다. mAP는 객체의 검출 여부와 검출 영역의 정확도를 함께 확인할 수 있는 평가 지표로 정면 카메라 영상에서 손 영역을 찾아서 지문자를 인식해야 하는 본 연구 목표의 성능을 확인하기에 적절한 평가 지표이다.

YOLO에서는 일반적으로 300 epoch으로 학습하는 것을 권장하고 있지만, 본 논문에서는 최적의 모델을 찾기 위하여 '100', '300', '400', '500', '1000'과 같이 5 가지 epoch으로 모델을 학습하여 성능을 비교하였다.

Table 4는 3 개의 학습 데이터셋에 대해 각 epoch 별 mAP50-95의 결과를 나타내고 있다. 실험 결과 400 epoch에서 가장 좋은 성능을 나타내었고, 이것을 본 논문의 최적 모델로 선정하였다.

400 epoch에서 데이터셋 별로 결과를 보면, 지문자 당 300개씩 총 9,300개의 이미지로 구성된 dataset_03을 사용한 경우가 0.929로 가장 좋은 결과를 내는 모델임을 확인하였다.

Table 4. Comparison of mAP50-95 by epoch

type	epoch 100	epoch 300	epoch 400	epoch 500	epoch 1000
dataset_01	0.875	0.896	0.899	0.896	0.892
dataset_02	0.904	0.917	0.919	0.917	0.917
dataset_03	0.925	0.928	0.929	0.928	0.928

4.3 지문자 인식 모델의 정답률 비교

YOLOv8n을 이용한 제안하는 모델과 SVM 기반의 기존 모델의 F1-Score 성능 비교 결과는 Table 5와 같다. Model_01에서 YOLOv8n은 0.981, SVM은 0.970의 점수를 보여 YOLOv8n이 1.1% 더 높은 성능을 보였다. Model_02에서는 YOLOv8n이 0.988, SVM이 0.966을 기록하여 YOLOv8n이 2.3%의 개선율을 보였다. Model_03에서 YOLOv8n은 0.990, SVM은 0.972의 점수를 보여 YOLOv8n이 1.9% 더 우수한 성능을 보였다. 모든 모델에

서 YOLOv8n이 SVM보다 높은 F1-Score를 기록하였으며, 특히 Model_02에서 가장 큰 성능 차이를 보였다.

Table 5. Performance Analysis(F1-Score)

Type	SVM	YOLOv8n	Improvement Rate
Model_01	0.970	0.981	1.1%
Model_02	0.966	0.988	2.3%
Model_03	0.972	0.990	1.9%

지문자 인식 모델의 31 개 전체 지문자 대한 정답률 분석 결과는 Table 6과 같다. Model_01에서 YOLOv8n은 99.4%, SVM은 97.1%의 정답률을 보여 YOLOv8n이 2.4% 더 높은 성능을 나타냈다. Model_02에서는 YOLOv8n이 99.7%, SVM이 96.8%의 정답률을 기록하여 YOLOv8n이 3.0%의 개선율을 보였다. Model_03에서 YOLOv8n은 99.9%, SVM은 97.2%의 정답률을 보여 YOLOv8n이 2.8% 더 우수한 성능을 보였다. 모든 모델에서 YOLOv8n이 SVM보다 높은 정답률을 기록하였으며, 정답 정확률에서도 Model_02에서 가장 큰 성능 차이를 보였다.

F1-Score와 정답률을 통해 제안하는 YOLOv8n 기반 모델이 기존 SVM 기반 모델보다 전반적으로 우수한 인식 성능을 나타내는 것을 확인하였다.

Table 6. Percentage of Correct Answers in the Fingerspelling Recognition Model

Type	SVM	YOLOv8n	Improvement Rate
Model_01	97.1%	99.4%	2.4%
Model_02	96.8%	99.7%	3.0%
Model_03	97.2%	99.9%	2.8%

한글 지문자 인식에서의 오인식에 관한 연구[16]에서 손가락 끝이 카메라 정면을 향하는 특정 지문자 'ㅓ','ㅔ','ㅕ','ㅖ'에서 오인식률이 높았다. 그래서 이 4 개의 문자에 대한 정답률을 따로 비교해 보았다. 분석 결과는 Table 7과 같다.

Model_01에서 YOLOv8n은 97.5%, SVM은 90.0%의 정답률을 보여 YOLOv8n이 8.3% 더 높은 성능을 나타냈다. Model_02에서는 YOLOv8n이 97.5%, SVM이 77.5%의 정답률을 기록하여 YOLOv8n이 25.8%의 큰 개선율을 보였다. Model_03에서 YOLOv8n은 99.2%, SVM은 80.8%의 정답률을 달성하여 YOLOv8n이 22.8% 더 우수한 성능을 보였다. 모든 모델에서 YOLOv8n이 SVM보다 높은 정답률을 기록하였으며, 특히 Model_02와 Model_03에서 매우 큰 성능 차이를 보였다. 또한 정답 정확률이 최소 97% 이상의 값을 나타내어 다른 문자에 대한 정확률과 비교하여 떨어지지 않는 성능을 보임으로써 제안하는 지문자 인식 모델이 전체 지문자 인식 알고리즘으로 활용될 수 있음을 보였다.

Table 7. Percentage of Correct Answers for Fingerspelling 'ㅓ', 'ㅔ', 'ㅕ', and 'ㅖ'

Type	YOLOv8n	SVM	Improvement Rate
Model_01	97.5%	90.0%	8.3%
Model_02	97.5%	77.5%	25.8%
Model_03	99.2%	80.8%	22.8%

5. 결론

본 논문에서는 YOLOv8n을 이용한 한글 지문자 인식 모델을 제안하고, 모델 학습에 사용할 데이터셋을 구축하기 위하여 미디어파이프 Hands Landmark를 이용한 자동 어노테이션 알고리즘을 제안하였다.

그리고 데이터셋 크기에 따른 성능을 비교하여 향후 지속적인 한글 지문자 연구를 지속하는데 있어서의 데이터셋 크기를 설정하는데 도움이 되도록 하였다.

제안하는 YOLOv8n 모델은 400 epoch으로 학습했을 때 성능 지표 mAP50-95에서 최고의 성능을 보였으며, 무작위로 추출한 9,300개 이미지 중 6,510개 이미지로 학습한 Model_03에서 0.929로 이보다 적은 이미지를 사용한 Model_01, Model_02보다 좋은 성능을 보였다. 이는 더 많은 학습 데이터를 사용할수록 모델의 성능이 향상됨을 시사한다.

또한 제안하는 모델의 성능을 검증하기 위하여 Landmark 정보를 이용한 기존 SVM 모델의 실험 결과와 비교하였다. F1-Score를 기준으로 한 YOLOv8n과 SVM 모델의 성능 비교에서 제안하는 YOLOv8n이 실험한 모든 데이터셋에서 일관되게 우수한 결과를 보였다.

이러한 결과는 학습 데이터의 양이 증가함에 따라 모델의 성능이 향상되는 것으로 판단되며 더 많은 고품질 데이터를 확보하여 학습에 활용한다면 성능을 더욱 개선할 수 있을 것으로 기대된다.

한글 지문자 전체 31개 지문자에 대한 정답률 분석에서도 제안하는 모델이 모든 테스트 모델에서 기존 모델보다 우수한 성능을 보였다.

주목할 만한 점은 선행연구에서 오인식률이 높았던 특정 지문자에 대한 성능 비교에서 제안하는 모델이 기존 모델보다 월등히 우수한 결과를 보였다는 것이다. 이 특정 지문자들에 대해 기존 모델보다 제안하는 모델이 8.3%에서 25.8%까지의 큰 성능 향상을 보였다.

이러한 결과는 YOLOv8n 기반 모델이 전반적인 지문자 인식뿐만 아니라, 특히 오인식 가능성이 높은 지문자에 대해서도 뛰어난 인식 능력을 가지고 있음을 시사한다. 따라서 본 논문에서 제안하는 YOLOv8n 기반 모델은 한글 지문자 31개 전체에 대해 97% 이상의 정답 정확도를 나타냄으로써 지문자 교육에 활용할 수 있는 가능성을 충분히 보여주었다.

본 연구 결과는 향후 더 정확하고 신뢰성 있는 한글 지문자 인식 시스템 개발에 중요한 기여를 할 것으로 기대된다.

참고문헌

[1] Lee. J. (2020). *Report on the 2020 Korean Sign language utilization survey results*. National Institute of Korean Language.

[2] Lee. C., Kim. J., Park. K., Jang. W., & Bien. Z. (1998). Implementation of Real-time Recognition System for Continuous Korean Sign Language (KSL) mixed with Korean Manual Alphabet (KMA). *The Institute of Electronics Engineers of Korea - C, 35*(6), 464-475.

[3] Kim. J., & Kang. E. (2022). Korean Finger Spelling Recognition Using Hand Landmarks. *The Journal of Korean Institute of Next Generation Computing, 18*(1), 81-91. https://doi.org/10.23019/kingpc.18.1.202202.008

[4] Karen E., & Jennifer A. (2012). Processing Orthographic Structure: Associations Between Print and Fingerspelling. *The Journal of Deaf Studies and Deaf Education, 17*(2), 194-204. https://doi.org/10.1093/deafed/enr051

[5] Park. J., Kim. Y., & Park. C. (2020). Real-Time Video and Data Analysis for Korean Sign Language Translation of Hearing Impaired People. *Korean Association of Computer Education Conference, 24*(2), 251-254. Seoul, Korea.

[6] Min. S., Oh. S., Kim. G., & Yoon. T. (2007). Optimize Data Glove-based System for Korean Finger Spelling Recognition. *Korean Institute of Information Scientists and Engineers, 34*(1), 237-241.

[7] Lee. S., Nam. J., Choi. J., Kim. K., & Kim. E. (2023). Glove-Type Sign Language Translator for Communication with the Hearing-impaired Person. *Journal of Rehabilitation Welfare Engineering & Assistive Technology, 17*(1), 35-40. https://doi.org/10.21288/resko.2023.17.1.35

[8] Kim. K., & Woo. Y. (2008). Real-Time Video and Data Analysis for Korean Sign Language Translation of Hearing Impaired People. *Journal of information and communication convergence engineering, 12*(6), 1101-1106

[9] Kim. N. (2011). A Development of the Next-generation Interface System Based on the Finger Gesture Recognizing in Use of Image Process Techniques. *Journal of information and communication convergence engineering, 15*(4), 935-942.

[10] Lee. D., Kim. M., Kim. N., & Choi. G. (2023). Implementation of Real-Time Sign Language AI Translation Program. *Journal of Digital Contents Society, 24*(10), 2585-2591 https://doi.org/10.9728/dcs.2023.24.10.2585

[11] Kim. M., Yoo. M., Jang. J., Choi. J., Na. G., Kim. M., & Na. J. (2022). Implementation of deep learning-based sign language acquisition online platform. *Journal of Digital Contents Society, 23*(11), 2147-2157. https://doi.org/10.9728/dcs.2022.23.11.2147

[12] Joseph Nelson. (2024). What is YOLO? The Ultimate Guide. https://blog.Roboflow.com/guide-to-yolo-models/

[13] Selcuk, B., & Serif, T. (2023). A comparison of yolov5 and yolov8 in the context of Mobile ui detection. *In International Conference on Mobile Web and Intelligent Information Systems*. 161-174. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-39764-6_11

[14] Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2023). Real-time flying object detection with YOLOv8. arXiv preprint. https://doi.org/10.48550/arXiv.2305.09972

[15] Cortes Corinna & Vapnik Vladimir. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

[16] Kim. J., & Kang. E., (2023). A Study on False Recognition in Korean Finger Spelling Recognition. *Korean Association of Computer Education Conference, 27*(1), 371-373. Yeosu, Korea.

[17] scikit-learn. https://scikit-learn.org

[18] MediaPipe, https://MediaPipe.readthedocs.io/en/latest/solutions/hands.html

[19] YOLO Datasets, https://docs.ultralytics.com/datasets/

[20] Roboflow ASL Dataset, https://public.Roboflow.com/object-detection/american-sign-language-letters/1

[21] Labelimg, https://github.com/HumanSignal/labelImg

김진영

· 2004년 순천대학교 정보통신공학과 (공학사)
· 2007년 순천대학교 컴퓨터교육전공 (교육학석사)
· 2011년 순천대학교 과학정보융합학과 (박사수료)
· 2023년 ~ 현재 순천대학교 컴퓨터교육과 시간강사

관심분야 : 인공지능, 컴퓨터비전, IoT

jykim@scnu..ac.kr

강의성

· 1991년 고려대학교 전자전산공학과 (공학사)
· 1995년 고려대학교 전자공학과 (공학석사)
· 1999년 고려대학교 전자공학과 (공학박사)
· 2001년 ~ 현재 순천대학교 컴퓨터교육과 교수

관심분야 : 영상처리, 신호처리, 인공지능, 컴퓨터교육

magasa@scnu.ac.kr

장문수

· 1992년 고려대학교 전자전산공학과 (공학사)
· 1994년 고려대학교 전자공학과 (공학석사)
· 2001년 동경공업대학 지능시스템전공 (공학박사)
· 2000년 ~ 2003년 한국전자통신연구원 선임연구원
· 2003년 ~ 현재 서경대학교 소프트웨어학과 부교수

관심분야 : 인공지능, 언어이해, 영상처리

cosmos@skuniv.ac.kr