



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

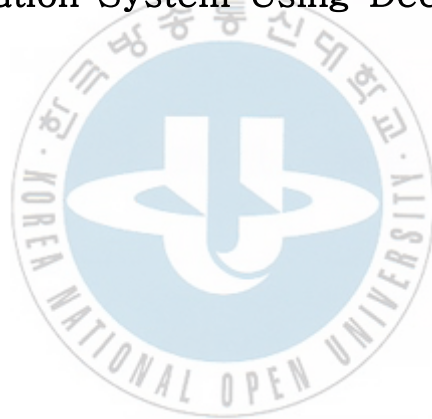
이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

# 딥러닝을 이용한 한국수화언어 통역 시스템 설계 및 구현

Design and Implementation of Korean Sign Language  
Interpretation System Using Deep Learning



한국방송통신대학교 대학원

정 보 과 학 과

구 민 재

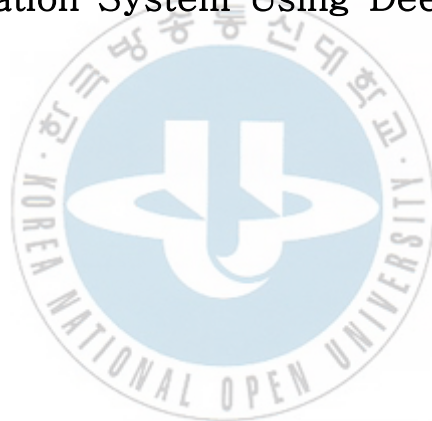
2019년



석사학위논문

# 딥러닝을 이용한 한국수화언어 통역 시스템 설계 및 구현

Design and Implementation of Korean Sign Language  
Interpretation System Using Deep Learning



한국방송통신대학교 대학원

정보과학과

구민재

2019년

# 딥러닝을 이용한 한국수화언어 통역 시스템 설계 및 구현

Design and Implementation of Korean Sign Language  
Interpretation System Using Deep Learning

지도교수 이 병 래

이 논문을 석사학위 논문으로 제출함

한국방송통신대학교 대학원

정 보 과 학 과

구 민 재

2019년 06월

구 민 재의  
석사학위 논문을 인준함

 심사위원장 \_\_\_\_\_인

심사위원 \_\_\_\_\_인

심사위원 \_\_\_\_\_인

한국방송통신대학교 대학원

2019년 06월

# 논 문 요 약

## 딥러닝을 이용한 한국수화언어 통역 시스템 설계 및 구현

구 민 재

한국방송통신대학교 대학원  
정 보 과 학 과  
(지도교수 이병래)

한국수화언어는 한국수화언어법으로 국어와 동등한 지위의 언어이고 청각장애인의 언어권을 신장하고 삶의 질을 향상하는 것을 목적으로 한다고 규정되어 있다. 한국수화언어 사용에 따른 소통의 어려움이 없어야 하지만, 한국수화언어를 사용하는 농인(聾人) 및 청인(聽人)과 한국어를 사용하는 일반인의 소통을 위해서는 일반이 한국수화언어를 배워야 하는 어려움이 있다. 이러한 소통의 문제 개선하기 위해 한국어를 사용하는 일반인이 한국수화언어의 의미를 이해할 수 있도록 딥러닝을 이용한 한국수화언어 통역 시스템을 제안하고자 한다.

한국수화언어 통역 시스템은 한국수화언어 특징 추출 모듈과 한국수화언어 통역 모듈로 구성된다. 한국수화언어 특징 추출 모듈은 객체 검출 모듈로, 손과 손가락 모양인 수형의 종류 및 얼굴 종류의 확률과 위치 정보를 추출한다. 한국수화언어 특징 추출 모듈은 한국수화언어 특징을 71개의 클래스로 분류하여 검출 할

수 있으며, 연산량은 0.615 BFLOPS로, CPU 연산을 사용한 객체 검출 소요 시간은 75ms이다. 한국수화언어 특징 추출 모듈 정확도 성능은 mAP(mean Average Precision) 94.02%, F1-score 87%, 평균 IoU 61.24%이다. 마지막으로 한국수화언어 특징 추출 모듈 크기는 6.92MB로 경량화된 모듈이다.

한국수화언어 통역 모듈은 한국수화언어 특징 추출 모듈을 사용하여 추출된 정보를 바탕으로 한국수화언어 중 자연수화언어 동작(단어)을 통역하는 모듈이다. 한국수화언어 통역 모듈은 한국수화언어 일상생활 대표 표제어 10개를 대상으로 제작되었고, 한국수화언어 통역 모듈의 정확도는 훈련에 사용되지 않은 평가 데이터 1,760개로 평가 시 95.85%로 나타났다. 한국수화언어 통역 App은 멀티스레드 기법을 적용하면 CPU 연산 사용 시 한국수화언어 통역에서 출력까지 25ms 소요되며, 28 FPS의 성능을 나타냈다. 이는 실시간으로 한국수화언어 통역이 가능한 성능이라 평가된다.

딥러닝을 이용한 한국수화언어 통역 시스템의 연구 가치는 다음과 같다. 첫째, 한국수화언어 통역 시스템은 실시간 통역 성능을 가진다. 둘째, 특수장비가 필요 없고 GPU와 같은 고성능 하드웨어를 필요로 하지 않는다. 셋째, 한국수화언어 훈련 데이터 수집이 용이하다. 마지막으로 PC 및 증강/혼합현실용 HMD와 같은 웨어러블 디바이스와 모바일 디바이스에 포팅이 가능하다.



## - 목 차 -

<b>제1장 서론</b> .....	<b>1</b>
1.1. 연구배경 및 필요성 .....	1
1.2. 연구목적 .....	2
1.3. 논문의 구성 .....	3
 <b>제2장 이론적 배경</b> .....	 <b>4</b>
2.1. 수화언어 .....	4
2.1.1. 수화언어의 특징 .....	4
2.1.2. 언어의 구성과 자질 .....	4
2.1.3. 수화언어 구성 .....	6
2.1.4. 수화언어 인식 연구 .....	10
2.2. 딥러닝 .....	15
2.2.1. 딥러닝의 종류 .....	15
2.2.2. 딥러닝을 이용한 객체 인식 .....	17
2.2.3. 딥러닝 모델 .....	18
2.3. 낮은 성능의 웨어러블 장치를 위한 손 검출 모델 .....	23
 <b>제3장 제안 방법</b> .....	 <b>27</b>
3.1. 한국수화언어 통역 시스템 개요 .....	27
3.2. 한국수화언어 데이터셋 .....	29
3.3. 한국수화언어 특징 추출 모듈 .....	36
3.4. 한국수화언어 통역 모듈 .....	41
3.5. 한국수화언어 통역 App .....	48

제4장 실험 .....	51
4.1. 한국수화언어 특징 추출 모듈 실험 및 평가 .....	51
4.2. 한국수화언어 통역 모듈 실험 및 평가 .....	52
4.3. 한국수화언어 통역 App 실험 및 평가 .....	53
4.4. 한국수화언어 통역 시스템 고찰 .....	55
 제5장 결론 .....	 57
 참고문헌 .....	 58
외국어초록 .....	61
부록 .....	63



## - 그림 목 차 -

<그림 1-1> 한국수화언어 통역 시스템의 증강현실 HMD 적용 예시 .....	3
<그림 2-1> 수화언어의 구성[8] .....	6
<그림 2-2> 영어 지문자 .....	7
<그림 2-3> 한국어 지문자 .....	8
<그림 2-4> 한국수화언어 수형 분류표[10] .....	9
<그림 2-5> 데이터 클러브 .....	11
<그림 2-6> 키넥트 XboX360 .....	12
<그림 2-7> 웨어러블 수화통역기 .....	12
<그림 2-8> 수화언어 인식 입력 이미지 사이즈 .....	13
<그림 2-9> Structure of the Neural Network Model .....	13
<그림 2-10> R-CNN 파이프라인[18] .....	17
<그림 2-11> YOLO 모델 파이프라인[20] .....	19
<그림 2-12> YOLO 모델 아키텍처[20] .....	20
<그림 2-13> Fire Module[23] .....	21
<그림 2-14> 5-layer dense block[24] .....	22
<그림 2-15> 홀로렌즈[25] .....	23
<그림 2-16> HoloLensHandNet 모델 아키텍처[27] .....	25
<그림 2-17> 홀로렌즈 손 검출기 앱 아키텍처[27] .....	27
<그림 3-1> 한국수화언어 통역 시스템 구성안 .....	28
<그림 3-2> 한국수화언어 특징 추출 데이터셋 .....	31
<그림 3-3> 한국수어사전 수화언어 영상[29] .....	32
<그림 3-4> 한국수화언어 일상생활 대표 표제어 데이터셋 .....	36
<그림 3-5> YOLO 모델 아키텍처[20] .....	37
<그림 3-6> 한국수화언어 특징 추출 모델 실행 화면 .....	39
<그림 3-7> 한국수화언어 특징 추출 데이터 .....	40

<그림 3-8> 한국수화언어 특징맵 예시 .....	43
<그림 3-9> 한국수화언어 통역 DesneNet 모듈 .....	45
<그림 3-10> 한국수화언어 통역 App 구동 프로세스 블록도 .....	48
<그림 3-11> 한국수화언어 통역 App 구현 .....	50
<그림 4-1> 한국수화언어 통역 App 구동 화면 .....	54



## - 표 목 차 -

<표 2-1> 딥러닝 알고리즘 종류와 활용분야[17] .....	16
<표 2-2> tiny-yolo 모델 네트워크 파라미터 .....	20
<표 2-3> 홀로렌즈 하드웨어 사양[26] .....	24
<표 3-1> 한국수화언어 수형코드 체계 .....	29
<표 3-2> 한국수화언어 얼굴코드 체계 .....	30
<표 3-3> 한국수화언어 일상생활 대표 표제어 .....	33
<표 3-4> 한국수화언어 일상생활 대표 표제어 데이터셋 리스트 .....	35
<표 3-5> KSL-yolov3-tiny 모듈 파라미터 .....	38
<표 3-6> KSL-HoloLensNet 모듈 파라미터 .....	38
<표 3-7> KSL_NLP_dataset_yolov3.json 예시 .....	42
<표 3-8> KSL_NLP_dataset.json 예시 .....	43
<표 3-9> 한국수화언어 통역 모듈 사용 파라미터 .....	44
<표 3-10> 한국수화언어 통역 모듈 Train batch 결과 예시 .....	46
<표 3-11> 한국수화언어 통역 모듈 검증 예시 .....	46
<표 3-12> 한국수화언어 통역 서버 적용 예시 .....	47
<표 4-1> 한국수화언어 특징 추출 모듈 평가 .....	52
<표 4-2> 한국수화언어 통역 모듈 평가 .....	53
<표 4-3> 한국수화언어 통역 App 평가 .....	54

# 제1장 서론

## 1.1. 연구배경 및 필요성

2016년 제정된 한국수화언어법 제1조에서는 “한국수화언어가 국어와 동등한 자격의 공식적인 언어임을 밝히고 그 발전과 보급의 기반을 마련함으로써 청각 장애로 인하여 일상생활에 제약을 받고 있는 청각장애인의 언어권을 신장하고 삶의 질을 향상하는 것을 목적으로 한다”라고 규정함으로써 언어로서의 한국수화언어의 언어권을 확립하였다[1].

수화언어의 명칭으로 ‘수화(手話)’와 ‘수어(手語)’가 존재한다. 언어의 관점으로 수어라 표현하지만, 관습적으로 수화로 많이 사용된다[2-3]. 이에 본 논문에서는 수어를 수화언어(Sign Language)로 한국수어를 한국수화언어(Korea Sign Language, KSL)로 사용한다.

한국수화언어를 사용하는 농인(聾人) 및 청인(聽人)과 한국어를 사용하는 일반인의 소통을 위해서는 한국수화언어를 배워야 하는 어려움이 있다. 이러한 소통의 문제 개선하기 위해 한국어를 사용하는 일반인이 한국수화언어의 의미를 이해할 수 있도록 한국수화언어 통역 연구가 필요하다.

한국수화언어 통역 연구는 언어적 관점에서 한국수화언어와 외국 수화언어가 다르기 때문에 외국의 수화언어 연구를 한국수화언어에 적용하는 것은 어려움이 있다. 또한, 한국수화언어에도 자연수화언어, 문법수화언어, 지문자 등으로 연구가 나뉘어 한국수화언어 통역 연구가 필요하다.

기존 한국수화언어 통역 연구는 수화언어 사용자가 특수장비를 착용하거나, 데스크 카메라 이용한 방법 등의 연구가 진행되었다. 이는 수화언어 사용자가 장비를 착용해야 하는 번거로움과 공간 제약 등 문제로 보편적 적용에 어려움이 있다[4-5]. 이러한 문제를 개선하려면 장비 착용의 번거로움과 공간 제약 문제에서 자유로운 한국수화언어 통역 연구가 필요하다.

한국수화언어 통역 연구에서 특수장비나 공간 제약 문제에서 자유롭기 위해서는 한국수화언어의 특징 검출을 영상 분석 및 검출 방법으로해야 보편적 적용이 가능하다. 영상 분석 및 검출 방법은 딥러닝 알고리즘의 발달로 딥러닝을 이용한 객체 검출 연구에서 좋은 성능을 내고 있다[6].

딥러닝 알고리즘은 컴퓨터 비전 문제를 해결하는 데 유용한 도구이며, 딥러닝 기술을 사용하기 위해서는 고성능의 GPU(Graphics Processing Unit)가 필요하고, 딥러닝을 이용한 검출 연구에는 GPGPU(General-Purpose computing on GPU) 기술이 사용되고 있다[6]. 그러나 증강/혼합현실용 HMD(Head Mounted Display)와 같은 웨어러블 플랫폼에서는 하드웨어의 무게, 크기, 배터리의 성능 등의 이유로 고성능의 하드웨어 자원을 활용할 수 없는 것이 현실이며, 증강/혼합현실용 HMD와 같은 웨어러블 플랫폼에서 구동 가능한 경량화된 딥러닝 모듈 연구가 필요하다.

## 1.2. 연구목적

본 논문에서는 딥러닝을 이용한 한국수화언어 통역 시스템을 제시하고자 한다. 한국수화언어 통역 시스템은 한국수화언어 특징 추출 모듈과 한국수화언어 통역 모듈로 구성된다.

한국수화언어 특징 추출 모듈은 객체 검출 모듈로 손과 손가락 모양인 수형의 종류 및 얼굴 종류의 확률과 위치 정보를 추출하는 모듈이다. 한국수화언어 특징 추출 모듈은 증강/혼합현실 HMD와 같은 웨어러블 플랫폼에서 실시간 구동을 할 수 있을 만큼 빠르고 경량화된 모듈이어야한다.

한국수화언어 통역 모듈은 한국수화언어 특징 추출 모듈을 사용하여 추출된 정보를 바탕으로 한국수화언어 중 자연수화언어 동작(단어)을 통역하는 모듈이다.

본 논문에서는 한국어를 사용하는 일반 사용자가 한국수화언어를 사용하는 사용자의 자연수화언어 동작(단어)을 통역하여 한국수화언어의 의미를 이해할 수 있도록 도와주는 한국수화언어 통역 시스템을 제안한다. 한국수화언어 통역 시스템은 <그림 1-1>과 같이 증강현실 HMD에 적용될 수 있다.



<그림 1-1> 한국수화언어 통역 시스템의 증강현실 HMD 적용 예시

세부 목표는 다음과 같다.

- ① 한국수화언어 특징 추출 모듈 제시
- ② 한국수화언어 통역 모듈 제시
- ③ 한국수화언어 통역 시스템 제시

### 1.3. 논문의 구성

본 논문 구성은 다음과 같다. 제2장은 관련 연구로 수화언어, 딥러닝, 낮은 성능의 웨어러블 장치를 위한 손 검출 모델의 이론적 배경을 설명한다. 제3장에서는 한국수화언어 통역 시스템 제안 방법에 대하여 기술한다. 제4장은 한국수화언어 특징 추출 모듈, 한국수화언어 통역 모듈, 한국수화언어 통역 App의 실험 및 평가를 진행하고, 한국수화언어 통역 시스템을 고찰한다. 제5장에서는 논문 결론을 기술하고, 향후 연구 방향과 한국수화언어 통역 시스템 활용 방법에 대해 언급한다.



## 제2장 이론적 배경

### 2.1. 수화언어

수화언어(Sign Language)란 손, 팔을 포함한 신체를 이용하여 의사를 전달하는 시각적 동작 언어로 다른 언어와 같이 언어의 음운론, 형태론, 통사론, 사회성, 역사성이 존재한다. 수화언어의 특징 및 언어의 구성과 자질, 수화언어의 구성을 살펴보고, 수화언어 인식 연구를 통하여 연구 방법을 살펴보고자 한다.

#### 2.1.1. 수화언어의 특징

수화언어는 의사를 표현할 시 손, 팔의 위치에 의해 의미가 변할 수 있고 동작의 속도를 변화시키거나 표정을 함께 사용함으로써 감정의 표현이 가능하다. 또한, 국가, 문화권, 언어권에 따라 다른 형태의 수화언어를 사용한다.

수어학은 자연수화언어를 과학적으로 연구하는 학문이다. 음성언어 연구에서 자연언어를 대상으로하는 것처럼 수어학의 연구 대상은 인공수어가 아닌, 농인 공동체에 의해 자연발생적으로 생겨나고 후대에게 전달되는 자연수화언어이다. 수화언어는 농인이 사용하는 비음성적 언어이다. 주변 청인 공동체의 음성언어에 근거하지 않으며, 역사적으로 농인 집단에서 생성되고 발달한 자연언어이다[7].

#### 2.1.2. 언어의 구성과 자질

한국수화언어가 언어임을 알기 위해서는 언어의 구성 자질을 살펴보고, 한국수화언어가 여기에 해당하는지 보면 알 수 있다[7].

① 음성-청각 채널 : 언어는 성도를 통해 산출되고 신호는 귀를 통해 수용된다.

- ② 광범위한 전달과 방향적인 수신 소리는 그것의 근원에서부터 모든 방향으로 움직이고 장애물을 피하며, 수신자는 말 수리의 근원이 어디인지 찾아낼 수 있다.
- ③ 빠른 사라짐 : 음성적 신호는 더 많은 메시지를 위해 빠르게 사라진다.
- ④ 교환성 : 언어 공동체의 구성원은 메시지의 발신자와 수신자가 될 수 있다.
- ⑤ 완전한 피드백 : 화자는 즉각적으로 들을 수 있고 그리하여 피드백에 의해 메시지를 모니터할 수 있다.
- ⑥ 특수화 : 말하기 행위는 언어의 의사소통 기능을 위해 전문화된다. 말하기는 어떤 부가적인 생리적 기능을 하지 않으며, 화자는 말하는 동안 다른 활동을 수행하는데 자유롭다.
- ⑦ 의미성 : 단지 언어의 지시적 기능만 고려한다. 인간의 언어는 의사소통의 의미적 체계이다.
- ⑧ 자의성 : 신호-사물 관계는 자의적이지도 대상적이지도 않다.
- ⑨ 전위성 : 언어 기호는 시간과 공간에서 멀리 떨어진 사물을 언급할 수 있다.
- ⑩ 기만 : 우리는 거짓이거나 혹은 의미 없는 것을 말할 수 있다.
- ⑪ 재귀용법 : 언어는 언어에 대해 의사소통하기 위해 사용될 수 있다. 언어의 메타 언어적 기능 있다.
- ⑫ 전통 : 언어의 관습은 생식질(germ plasm)을 통해서가 아니라 가르침과 학습에 의해서 전해진다.
- ⑬ 학습 가능성 : 한 언어의 화자는 또 다른 언어를 배울 수 있다.
- ⑭ 불연속성 : 기호 레퍼토리는 불연속적이고 재발하는 단위로 구성된다. 언어적 요소는 더 혹은 덜이라는 점에서 단계적 차이가 없다.
- ⑮ 창조성 혹은 개방성 : 새로운 언어적 메시지는 자유롭고 쉽게 조어되고, 맥락에서 보통 이해된다. 새로운 메시지는 언어 기호들의 창조적이 결합으로 생성된다.

한국수화언어는 음성언어의 15가지 구성 자질 중에서 음성 자질을 뺀 12개 특성과 부합한다. 한국수화언어와 음성언어의 가장 뚜렷한 차이는 시각-제스처 채널이라는 특성이다. 따라서, 한국수화언어 통역 연구는 한국수화언어를 언어의 관점에서 접근해야 한다.

### 2.1.3. 수화언어 구성

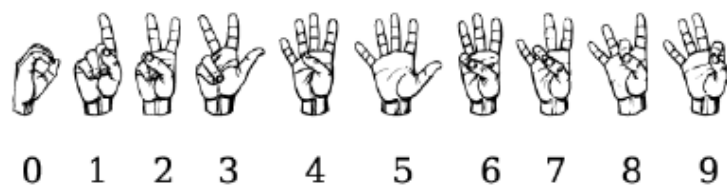
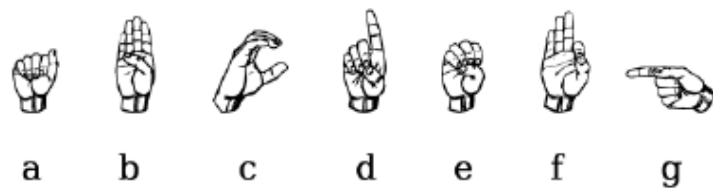
한국수화언어는 수형, 수위, 수동, 수향으로 구성되며 이를 ‘수어소’라 칭한다. 한국수화언어 단어는 수지 요소인 수형, 수위, 수동, 수향 뿐만 아니라 손을 제외한 수화언어 사용자의 얼굴 표정, 상체의 방향과 움직임과 같은 비수지 요소로 구성된다. 수어소는 음성언어 단어 의미를 구별하는 소리인 ‘음소’에 해당하고, 말의 뜻을 구별해주는 소리의 최소 단위를 음운이라고 하고 음운은 자음과 모음인 음소와 소리의 길이, 높낮이, 강세와 같은 음소를 포함한다[7]. 수화언어의 구성 요소를 <그림 2-1>과 같이 분류할 수 있다.

구성	설명
	수향
	수형
	수동
	수위
비수지 신호	눈썹 움직임, 시선의 방향, 입 모양과 혀의 움직임, 얼굴 기울기와 방향, 얼굴 표정, 어깨 움직임, 발의 움직임

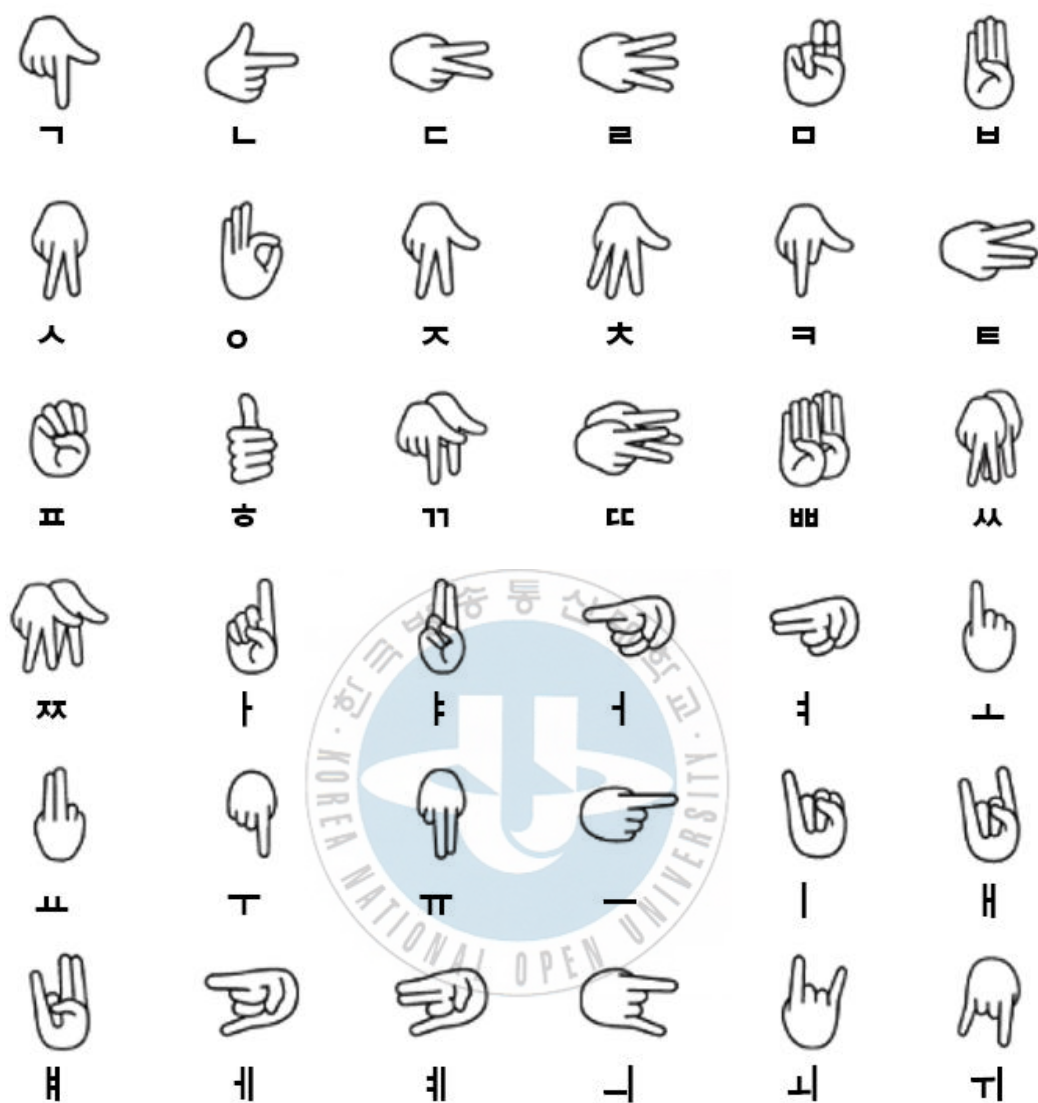
<그림 2-1> 수화언어의 구성[8]

음성인식에서도 음소 단위가 음성인식과 음성합성에 사용된다. 한국수화언어 특징 추출 모듈 개발을 위해서는 한국수화언어 특징 추출 방법이 필요하며, 수지 요소와 비수지 요소를 함께 포함할 수 있어야 한다.

수화언어로 지정되지 않거나 표현하기 어려운 단어를 표현할 때 지문자를 사용한다. 지문자는 국가별로 다양하게 나타날 수 있다. 미국수화언어(American Sign Language, ASL) 지문자 경우는 알파벳(‘a’, ..., ‘z’)에 해당하고 <그림 2-2>와 같다. 한국수화언어 지문자 경우 한국어 특성을 반영한 자음(‘ㄱ’, ..., ‘ㅎ’), 모음(‘ㅏ’, ..., ‘ㅣ’), 쌍자음(‘ㄲ’, ‘ㄸ’, ‘ㅃ’, ‘ㅆ’, ‘ㅈ’)에 해당하고 <그림 2-3>과 같다[9]. 이처럼 미국수화언어와 한국수화언어의 지문자는 언어권에 따라 다르게 구성된다. 국어 국립원의 “수형 기반 한국수어사전 구축” 연구를 통해 <그림 2-4>와 같이 손과 손가락의 모양인 수형을 65가지로 분류하였다[10].



<그림 2-2> 영어 지문자



<그림 2-3> 한국어 지문자

수형 사진													
코드	0000	0100	0200	0300	0400	0500	0600	0700	0800	0900	1000	1100	1200
수형 사진													
코드	0001	0101	0201	0301	0401	0501	0601	0701	0801	0901	1001	1101	
수형 사진													
코드	0002	0102	0202	0302	0402	0502	0602	0702	0802	0902			
수형 사진													
코드	0003		0203	0303	0403		0603	0703		0903			
수형 사진													
코드	0004		0204	0304			0604	0704		0904			
수형 사진													
코드	0005		0205	0305			0605	0705		0905			
수형 사진													
코드							0606	0706		0906			
수형 사진													
코드										0907	0911		
수형 사진													
코드										0908	0912		
수형 사진													
코드										0909	0913		
수형 사진													
코드										0910	0914		

<그림 2-4> 한국수화언어 수형 분류표[10]



#### 2.1.4. 수화언어 인식 연구

수화언어 인식 연구는 수화언어를 해석하고 재현하려는 것으로, 수화언어를 인식하는 문제, 한국어를 수화언어로 번역하는 문제, 한국어를 아바타를 이용하여 재현하는 문제, 수화언어를 한국어로 통역하는 문제, 수화언어를 음성언어로 변환하는 문제 등 다양한 연구 관점이 존재한다. 수화언어 인식 연구 종류에도 동작 기반의 자연수화언어 통역과 지문자 통역이 있다. 수화언어 특징을 추출하는 방법에는 특수장비 착용하는 방법 및 고정 장치를 이용한 방법, 이동형 장치를 이용한 방법이 있다. 본 논문에서는 딥러닝을 알고리즘을 이용하여 한국수화언어 자연수화언어 동작(단어)을 한국어 표제어로 통역하고자 한다. 이 문제 해결을 위해 선행 연구를 살펴보았다.

##### (1) 영상검출을 이용한 수화언어 인식

이병환 등은 PC 카메라의 영상검출을 통한 수화언어 인식 연구를 진행하였다[11]. RGB 모델 추출법과 HSI 모델 추출법을 사용하여 수화언어 특징을 추출하였으며 이를 40개의 단어를 인식하는 연구를 진행하였다. 연구의 한계점은 수화인식 시스템이 설치된 장소 이동의 어려운 공간 제약 한계점이 있다.

양희덕은 수지 신호와 비수지 신호를 결합하여 수화를 인식할 수 있는 계층적 구조로 구성된 방법을 제안하였다[12]. 첫 번째 단계에서는 계층적 Conditional Random Field(CRF)를 이용하여 수지 신호 후보 영역을 검출한다. 두 번째 단계에서는 BoostMap 임베딩을 이용하여 첫 번째 단계에서 검출된 수지 신호의 손 모양을 인식한다. 마지막 세 번째 단계에서는 Support Vector Machine(SVM)을 이용하여 비수지 신호 분석을 위한 얼굴 인식을 수행한다. 제안된 방법은 영어 지문자를 사용하는 청각장애인이 수행한 수화 문장에서 84%의 인식률을 보였다.

## (2) 데이터 글러브를 이용한 수화언어 인식

데이터 글러브를 이용한 수화언어 인식 연구는 <그림 2-5>와 같이 접촉식 센서 입력 방식으로 장갑 손가락 부분에 15개의 구부림 센서(flex sensor)가 손가락 구부림의 정도를 측정하고, 나침반, 가속도계, 자이로스코프가 제스처 모션의 공간적인 부분을 측정한다. 장갑의 마이크로컨트롤러로부터 입력된 센서 데이터는 블루투스를 통하여 모바일 디바이스로 전송된다. 이 데이터에 대해 패턴 인식을 함으로써 손과 손가락의 위치를 텍스트로 번역하는데, “Nice to meet you” 정도의 적은 어구 정도만 번역할 수 있다[13]. 수화언어 사용자가 특수장비인 데이터 글러브를 착용해야 인식할 수 있는 접촉형의 한계점이 있다.

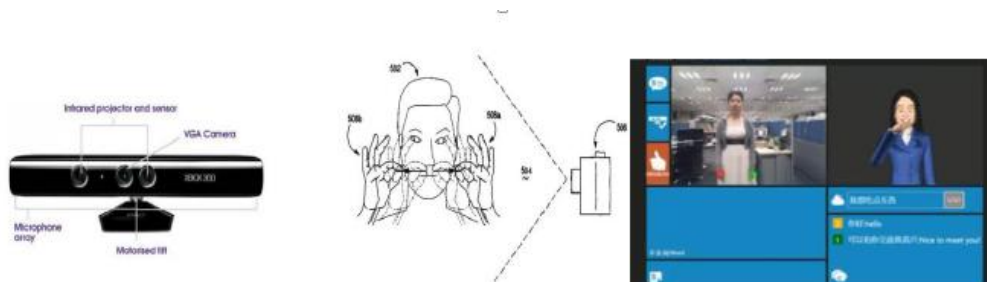


<그림 2-5> 데이터 글러브

## (3) 키넥트를 이용한 수화언어 인식

키넥트 XboX360은 비접촉 입력 방식으로 <그림 2-6>과 같이 웹 카메라 모션 컨트롤러를 이용하여 몸동작 전체를 3차원으로 인지하여 음성 방향 및 거리를 감지한다. 스켈레탈 맵핑으로 48개 신체 부위를 인지할 수 있으며 뼈, 관절, 팔, 작은 신체 부위인 손가락과 발가락의 인식이 가능하며 영어와 중국어 370개 단어만 사용할 수 있다[14]. 이 연구는 영어와 중국어의 수화언어를 대상으로 진행되어 한국수화언어를 사용하지 못한다. 또한, 얼굴 표정과 같은 비수지 요소의 정보를 포함할 수 없는 한계점과 공간 제약의 한계점이 있다.





<그림 2-6> 키넥트 XboX360

#### (4) 립모션을 이용한 수화언어 인식

<그림 2-7>과 같이 립모션이라는 제스처 인식 디바이스를 이용하여 수화언어를 텍스트 및 음성언어로 나타내는 연구 결과도 있다[15]. 수화언어를 표현에 있어 문자와 음성언어를 동시에 사용하여 나타냄으로써 사용자에게 수화언어 의미전달을 더 명확하게 해줄 수 있다. 또한, 립모션을 목걸이 형태의 웨어러블 디바이스로 제작함으로써 공간의 제약 없이 어디서든지 사용할 수 있게 하여 휴대성 및 편의성을 높였다. 이 연구는 특수장비인 립모션이 필요하며, 립모션은 손과 손가락의 제스처를 인식하는 디바이스이기 때문에 얼굴 표정과 같은 비지수 요소를 포함하지 못하는 한계점이 있다.

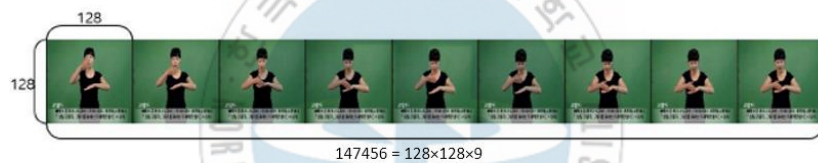


<그림 2-7> 웨어러블 수화통역기

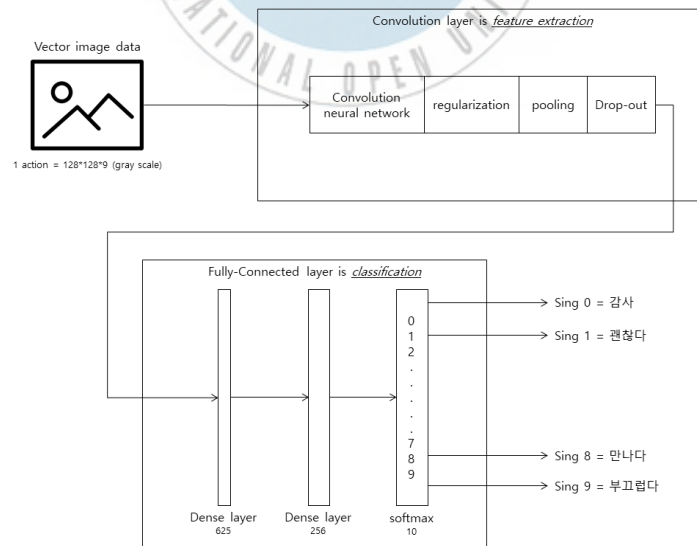
## (5) 딥러닝을 인용한 수화언어 인식

딥러닝을 수화언어 인식에 적용하는 사례도 있다. CNN 기법인 GoogLeNet 모델을 사용하여 실시간 미국수어(ASL) 지문자 인식을 수행하였는데, Top-5 Accuracy 0.9163의 성능을 보였다[16].

한국수화언어에 대해서도 CNN을 적용한 연구가 진행된 바 있다[9]. 한국수화언어 영상을 기반으로 <그림 2-8>과 같이 수화언어 특징을 나타낼 수 있는 9장(그레이 채널)의 프레임을 추출하였다. 이  $128 \times 128 \times 9$ 의 입력 데이터에 대해 <그림 2-9>의 CNN의 훈련을 진행하였다. 한국수화언어 표제어 10개 대상으로 실험한 결과 정확도 84.52%를 보였다. 이 연구에서 제안된 모델은 한국수화언어 자연수화언어의 인식은 가능하지만, 전처리 프로세스 문제로 인해 실시간 한국수화언어 통역 문제에 적용하기 어렵다는 한계가 있다.



<그림 2-8> 수화언어 인식 입력 이미지 사이즈



<그림 2-9> Structure of the Neural Network Model

## (6) 수화언어 인식 연구 고찰

수화언어의 인식 및 분류에 관한 기존 연구는 여러 디바이스와 알고리즘을 혼용하여 사용된다. 연구의 큰 흐름은 크게 수화언어 지문자(KSL는 모음과 자음, ASL는 알파벳)를 인식하는 연구와 자연수화언어 동작(단어)를 인식하는 연구로 구분할 수 있다.

수화언어 인식과 분류를 위한 연구에서 제안된 방법에는 디바이스를 이용한 방법, 영상처리 알고리즘을 이용한 방법, 인공신경망을 이용한 방법 등이 있다.

디바이스를 이용한 수화언어 인식 연구에서는 수화언어 사용자가 사람의 몸에 착용하는 형태의 디바이스를 사용한다. 대표적으로 데이터 글러브를 이용한 수화언어 인식 방법[13]과 손과 손가락의 동작을 인식할 수 있는 립모션 같은 웨어러블 장치를 이용하는 방법[15]이 있다. 이러한 디바이스를 이용하는 방법은 수화언어의 지문자와 자연수화언어 동작(단어)을 실시간으로 처리할 수 있는 장점이 있다. 하지만 디바이스 장치를 이용해야 입력할 수 있다는 한계점이 있다.

영상처리 알고리즘을 이용한 수화언어 인식 알고리즘에 관한 연구에서는 수화언어 지문자 인식을 중심으로 한 영상처리 기법을 적용한다. 수화언어 사용자의 지문자를 토대로, 전문가가 각각의 지문자를 구분하고 특징을 선택하여 구분할 수 있는 알고리즘을 제시하였다. 이러한 알고리즘의 특징 선택 및 추출은 머신러닝 알고리즘 중 SVM(Support Vector Machine) 기법이 사용되었다[12]. 이러한 방법은 분류를 위한 특징을 분류할 수 있는 전문가를 필요로 하며, 특징이 잘못 분류되었을 때 정확히 분류할 수 없는 문제점을 갖고 있다.

딥러닝을 이용한 방법은 데이터 수집 비용에 따라 고비용과 저비용으로 구분할 수 있다. 고비용이란 딥러닝에서 이용되는 데이터를 수집하면서 특수장비를 사용하는 방법을 말하며, 저비용이란 데이터의 수집을 키넥트와 같은 특수장비가 아닌 일반적인 카메라 기능으로 촬영이 가능한 것을 말한다. 고비용 장비로는 모션에너지 영상 MHV(Motion History Volume)를 추적 및 촬영할 수 있는 카메라가 있다. 저비용 인공신경망을 이용한 수화언어 분석은 카메라 장비에 대한 제약은 적지만 수화언어 특징 추출은 딥러닝 알고리즘에 의존하게 된다.

따라서, 본 논문에서 제안하는 한국수화언어 통역 시스템은 수화언어 사용자의 자연수화언어를 통역하는 시스템으로 공간과 장소, 장비의 제약에서 자유로운 시스템을 제안하고자 한다. 한국수화언어 통역 시스템의 알고리즘은 한국수화언어 특징 추출하기 위해 딥러닝 객체 검출 모듈을 사용하는데, 웨어러블 디바이스에서도 동작할 수 있는 빠른 성능의 모듈을 제안한다. 한국수화언어 통역 모듈은 데이터 수집에 있어 저비용의 딥러닝 알고리즘을 사용하여 한국수화언어를 통역하고 한다.

## 2.2. 딥러닝

딥러닝은 컴퓨터 비전 문제를 해결하는 데 유용한 도구이다. 본 연구에서는 딥러닝 알고리즘을 이용하여 한국수화언어 통역 시스템을 구현하려고 한다. 이를 위해 딥러닝의 종류를 알아보고 한국수화언어 특징 추출 모듈설계에 필요한 딥러닝을 이용한 객체 인식 연구를 살펴보고자 한다. 또한, 한국수화언어 통역 시스템에 적용 가능한 모델을 살펴본다. 딥러닝 기술을 사용하기 위해서는 고성능의 GPU가 필요하고 딥러닝을 이용한 객체 검출 연구에는 GPGPU 기술이 사용되고 있다[6]. 그렇지만 현재 활용 가능한 증강/혼합현실용 HMD와 같은 웨어러블 디바이스에서는 고성능 하드웨어 자원을 사용할 수 없다. 본 연구에서는 이러한 문제 해결한 한국수화언어 통역 시스템을 제안하고자 한다.

### 2.2.1. 딥러닝의 종류

딥러닝은 여러 비선형 변환기법의 조합을 통해 높은 수준의 추상화를 시도하는 기계학습 알고리즘의 집합으로 정의되며, 큰 틀에서 사람의 사고방식을 컴퓨터에 가르치는 기계학습의 한 분야라고 이야기할 수 있다. 어떤 데이터가 있을 때 이를 컴퓨터가 알 수 있는 형태로 표현하고 이를 학습에 적용하기 위해 많은 연구가 진행되었다. 이러한 노력의 결과로 합성곱 신경망(Convolution Neural Network, CNN), 심층신뢰 신경망(Deep Belief Network, DBN), 순환 신경망(Recurrent Neural Network, RNN)과

같은 다양한 딥러닝 기법들이 개발되었다. <표 2-1>은 이러한 모델들의 특성을 요약한 것이다. 딥러닝의 학습 모형은 지도학습(Supervised Learning)과 비지도학습(Unsupervised Learning) 모형으로 나뉜다. 본 논문에서는 합성곱 신경망 기반 지도학습을 사용하여 한국수화언어 특징 추출 문제와 한국수화언어 통역 문제를 해결하려 한다.

<표 2-1> 딥러닝 알고리즘 종류와 활용분야[17]

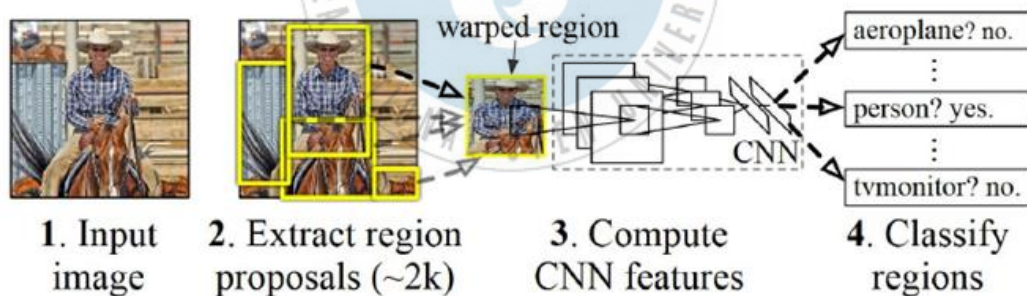
종류	개념	활용 분야
합성곱 신경망 (Convolution Neural Network, CNN)	최소한의 전처리를 사용하도록 설계된 다계층 퍼셉트론의 한 종류로, 하나 또는 여러개의 합성곱 계층과 그 위에 올려진 일반적인 인공 신경망 계층들로 이루어져 있으며, 가중치와 통합 계층(pooling layer)들을 추가 활용한다.	영상, 음성분석
심층신뢰 신경망 (Deep Belief Network, DBN)	기계학습에서 사용되는 그래프 생성 모형(generative graphical model)으로, 잠재변수(latent variable)의 다중계층으로 이루어진 심층 신경망이다. 계층 간에는 연결이 있지만, 계층 내의 유닛 간에는 연결이 없다.	영상, 음성, 자연어처리 등 전 분야
순환 신경망 (Recurrent Neural Network, RNN)	순환 신경망은 인공신경망을 구성하는 유닛 사이에 순환 연결이 존재한다. 순환 신경망은 앞먹임 신경망과 달리, 임의의 입력을 처리하기 위해 신경망 내부의 메모리를 활용할 수 있다.	필기체 인식, 자연어 처리

### 2.2.2. 딥러닝을 이용한 객체 인식

딥러닝을 이용한 객체 인식은 여러 객체 위치를 검출하는 문제와 검출된 객체의 클래스를 분류하는 문제를 포함한다. 이 문제를 해결할 수 있다면 기존의 영상 처리 알고리즘을 기반으로 구축된 응용 사례에서 매우 유용하게 사용될 수 있고, 수화언어처럼 손과 손가락을 검출하여 한국수화언어 통역 문제를 해결에 응용될 수 있다.

영상 처리의 객체 검출 연구는 저수준(low-level) 특징에 기반한 SIFT(Scale-Invariant Feature Transform), SURF(Speeded-Up Robust Feature), HOG(Histogram of Oriented Gradient)와 같은 방법들이 사용되었다[6].

R-CNN(Region-based Convolutional Neural Networks)[18]은 Selective Search 알고리즘을 기반으로 입력 영상 내 객체 후보 영역을 제안하고, 각 영역에 대하여 CNN을 거쳐 SVM 분류기를 통한 영역 내 객체 패턴의 클래스를 분류한다. 이 CNN은 AlexNet 모델을 기반으로 객체 클래스 분류에 대하여 선행으로 학습된 상태의 모델을 가져온다. R-CNN의 파이프라인은 <그림 2-10>과 같다.



<그림 2-10> R-CNN 파이프라인[18]

Fast R-CNN은 마지막 컨볼루션 특징 맵에서의 객체 후보 영역 검출 시 여러 스케일의 풀링을 수행하는 레이어 대신 ROI(Region of Interest) 풀링 레이어를 적용한다. Fast R-CNN 자체에서 CNN의 컨볼루션 레이어들에 대한 미세 조정이 가능하다. Fast R-CNN은 R-CNN 비하여 학습 과정은 약 2.7배, 테스트 과정은 7배 이상 개선되었고,



평균 인식률이 개선되었다[19].

YOLO(You Only Look Once) 네트워크는 최종 출력단에서 경계 상자 검출과 클래스 분류가 동시에 수행되도록 설계되었다. 이것은 대부분의 딥러닝 기반 객체 인식 및 검출 연구가 물체의 위치 및 영역 추정 프로세스와 클래스 분류 프로세스를 별도의 CNN 기반으로 진행하는 것과 대비된다. PASCAL VOC 2007과 2012를 함께 사용한 데이터셋에 대해 YOLO 네트워크를 적용한 결과 63.4%의 평균정밀도(mAP, mean average precision)를 보였으며, GPU를 사용할 경우 45 FPS(frames per second)의 빠른 검출 성능과 실시간성을 보였다[20].

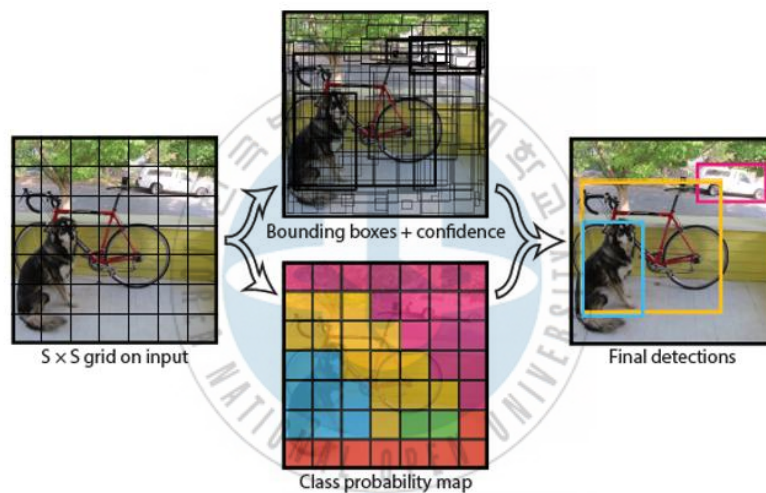
SSD(Single Shot multibox Detector)는 각 컨볼루션 레이어마다 YOLO의 최종 출력 텐서에 해당하는 특징 맵을 생성하도록 한다. 이를 통해 예측하는 경계 상자 후보가 YOLO보다 훨씬 많고 다양한 스케일의 객체 후보 경계 상자를 추정 가능해지므로 인식 정확도가 개선되는 효과를 보인다. 단일 네트워크 구조를 사용하면서 네트워크의 각 레이어가 객체 검출을 분할하여 수행하는 과정이기 때문에 검출 속도는 거의 유사한 수준을 유지할 수 있다[21].

### 2.2.3. 딥러닝 모델

한국수화언어 통역 시스템은 한국수화언어 특징 추출 모듈과 한국수화언어 통역 모듈로 구성되어 있다. 한국수화언어 특징 추출 모듈 제작을 위해서는 한국수화언어의 손과 손가락 모양인 수형을 빠르게 검출할 수 있는 성능이 필요하다. 증강/혼합현실용 HMD와 같은 웨어러블 디바이스의 동작을 위해서는 딥러닝 모델 연구가 필요하다. 본 연구에서는 실시간 객체 검출 모델인 YOLO[20]와 모델 압축 연구인 SqueezeNet[23]을 응용한 모델을 설계하여 한국수화언어 특징 추출 모듈을 제시하고자 한다. 한국수화언어 통역 모듈은 CNN 모델 중 DenseNet(Densely connected Convolutional Networks)[24]을 응용한 모델을 설계하여 제안하고자 한다.

## (1) YOLO(You Only Look Once)

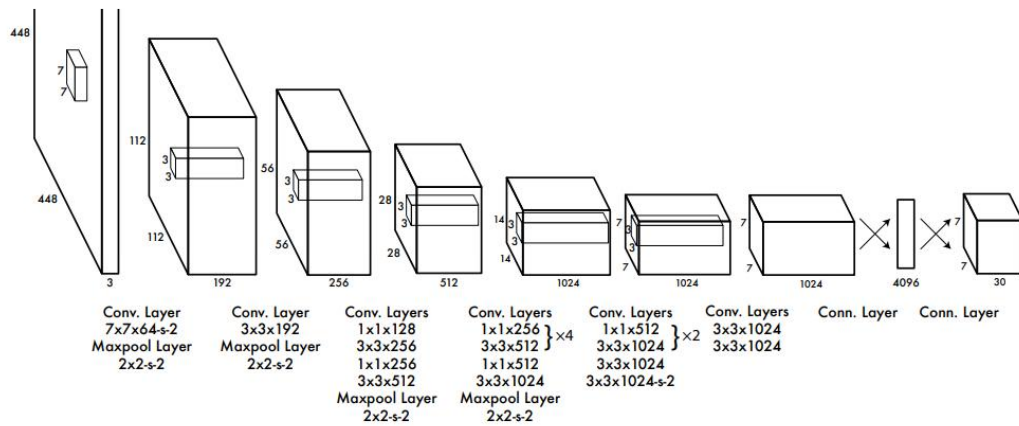
YOLO 모델[20]은 딥러닝 네트워크의 최종 출력의 경계 상자를 검출과 클래스 분류가 동시에 수행되도록 설계되었다. YOLO 모델은 객체 검출을 위해 이미지의 픽셀 경계 상자 좌표 및 클래스 확률을 회귀 문제로 정의하였다. <그림 2-11>과 같이 이미지를  $S \times S$  그리드로 나누고 각 그리드 셀은 예측된 박스와 비교하여 신뢰도 점수를 예측한다. 신뢰도 점수는 해당 경계 상자에 개체의 유무로 모델에 대한 신뢰도 점수를 반영한다.



<그림 2-11> YOLO 모델 파이프라인[20]

YOLO 모델 아키텍처는 <그림 2-12>과 같다. tiny-yolo는 YOLO 모델 중 가장 작은 연산량(5.41 BFLOPS)을 갖는 모델이며, GPU 사용 시 244 FPS의 빠른 성능을 낸다. tiny-yolo 모델의 네트워크 파라미터는 <표 2-2>와 같다. 한국수화언어 특징 검출 모듈에 기반 모델로 사용되었고, 이러한 성능을 바탕으로 증강/혼합현실용 HMD와 같은 웨어러블 디바이스에 적용을 위한 연구가 필요하다.





<그림 2-12> YOLO 모델 아키텍처[20]

<표 2-2> tiny-yolo 모델 네트워크 파라미터

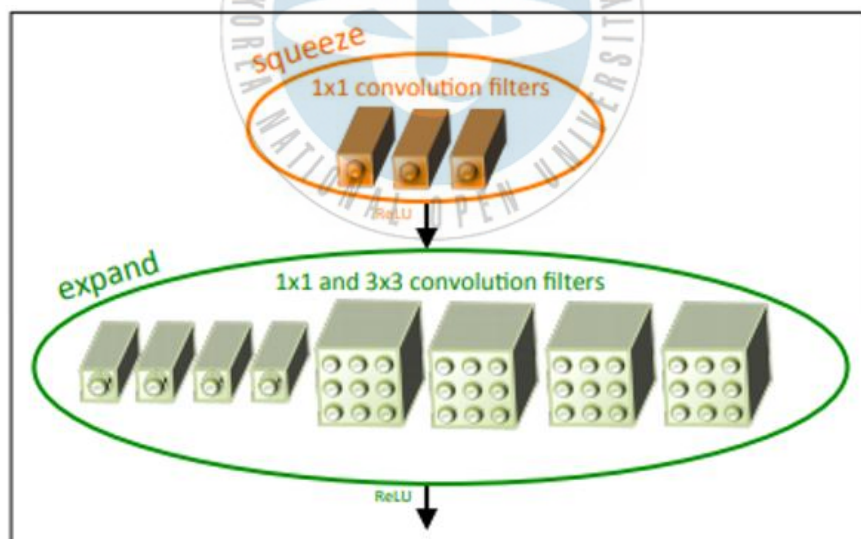
No	Layer	Filters	Size	Input	Output	BFLOPS
0	conv	16	3 x 3 / 1	416 x 416 x 3	416 x 416 x 16	0.150
1	max		2 x 2 / 2	416 x 416 x 16	208 x 208 x 16	0.003
2	conv	32	3 x 3 / 1	208 x 208 x 16	208 x 208 x 32	0.399
3	max		2 x 2 / 2	208 x 208 x 32	104 x 104 x 32	0.001
4	conv	64	3 x 3 / 1	104 x 104 x 32	104 x 104 x 64	0.399
5	max		2 x 2 / 2	104 x 104 x 64	52 x 52 x 64	0.001
6	conv	128	3 x 3 / 1	52 x 52 x 64	52 x 52 x 128	0.399
7	max		2 x 2 / 2	52 x 52 x 128	26 x 26 x 128	0.000
8	conv	256	3 x 3 / 1	26 x 26 x 128	26 x 26 x 256	0.399
9	max		2 x 2 / 2	26 x 26 x 256	13 x 13 x 256	0.000
10	conv	512	3 x 3 / 1	13 x 13 x 256	13 x 13 x 512	0.399
11	max		2 x 2 / 1	13 x 13 x 512	13 x 13 x 512	0.000
12	conv	1024	3 x 3 / 1	13 x 13 x 512	13 x 13 x 1024	1.595
13	conv	1024	3 x 3 / 1	13 x 13 x 1024	13 x 13 x 1024	3.190
14	conv	30	1 x 1 / 1	13 x 13 x 1024	13 x 13 x 30	0.010
15	detection					

YOLO 모델 개발 환경은 Darknet 프레임워크다. Darknet은 C와 CUDA로 작성된 오픈소스 신경망 프레임워크로 github를 통해 공개되었다. Darknet은 리눅스 환경과 윈도우 환경에서 구동할 수 있으며, 윈도우 OS 기반 DLL로 컴파일하여 사용할 수 있

다[20]. DarkNet은 윈도우 OS 기반 디바이스에 포팅하여 다양한 애플리케이션 개발에 응용할 수 있다.

## (2) SqueezeNet

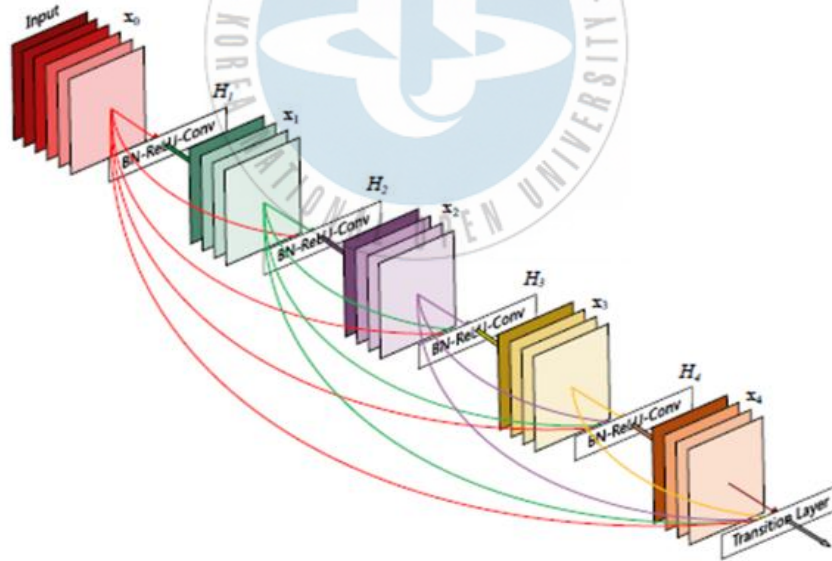
SqueezeNet은 AlexNet과 비슷한 성능을 내면서도 파라미터는 50배나 줄고, 모델 사이즈는 0.5MB밖에 안 되는 작은 규모의 모델이다[23]. SqueezeNet 네트워크 구조는 네트워크가 최대한 파라미터를 적게 갖지만 성능을 유지할 수 있도록 <그림 2-13>와 같은 Fire module을 제시하였다. Fire module의 핵심은  $3 \times 3$  컨볼루션 필터 대신하여  $1 \times 1$  컨볼루션 필터를 사용함으로써 파라미터를 줄여 연산 성능은 향상하고 정확도를 유지한다. 증강/혼합현실용 HMD와 같은 웨어러블 디바이스는 정확도는 유지하면서도 경량화된 모듈이 필요하다. 한국수화언어 특징 추출 모듈 딥러닝 네트워크 제작시 Fire module을 사용하면 정확도는 유지하면서 경량화된 모듈 제작이 가능하다.



<그림 2-13> Fire Module[23]

### (3) DenseNet(Dense connected Convolutional Networks)

DenseNet 모델은 CNN 기반 모델로 네트워크 구조는 Dense Block과 Transition Layer가 반복적으로 연결된 구조를 갖는다[24]. Dense Block은 합성곱층의 출력이 이후에 오는 모든 합성곱층들의 입력에 연결이 되도록 Dense connectivity 구조를 구축함으로써, 더 깊게 층을 쌓을 수 있도록 설계되었다. CNN 기반 모델에서는 합성곱층의 수가 늘어나면 전체 파라미터가 증가하여 훈련이 제대로 이루어지지 않는 현상이 발생한다. 이 현상을 방지하기 위해 합성곱층의 출력 특징맵의 수를 의미하는 성장률을 사용하여, 합성곱층을 이 증가하여 파라미터가 증가를 제한하였다. DenseNet 모델은 <그림 2-14>와 같이 파라미터 증가를 제한하였고, DenseNet 모델의 성능은 실험 결과 학습 데이터의 크기와 상관없이 좋은 성능 나타내었다. 한국수화언어 통역 모듈에 DenseNet 모델을 적용 연구를 진행하였다.



<그림 2-14> 5-layer dense block[24]

### 2.3. 낮은 성능의 웨어러블 장치를 위한 손 검출 모델

한국수화언어 통역 시스템은 HMD나 모바일 장치와 같은 휴대 가능한 장치에서 사용할 경우 유용성이 높다. 본 논문에서는 대표적인 증강/혼합현실용 HMD 장치인 홀로렌즈의 예를 통해 모바일 장치의 특성을 검토해 보고, 홀로렌즈를 기반으로 손 검출 모델을 구현한 HoloLensHandNet[27]의 사례를 살펴봄으로써 증강/혼합현실용 HMD와 같은 웨어러블 플랫폼에서 실시간 구동 가능한 한국수화언어 통역 시스템 적용 연구를 진행하고자 한다.

#### (1) 홀로렌즈

마이크로소프트 홀로렌즈는 <그림 2-15>와 같은 모습으로 2016년 상반기 개발자 버전이 출시되었으며, 개발자 버전에는 하드웨어인 홀로렌즈 글라스 본체와 소프트웨어 개발자 환경이 포함되어있다. 홀로렌즈는 자체적으로 윈도우 10 OS를 내장하고 있어 별도의 PC 없이 독립적으로 사용할 수 있으며 중앙처리장치(CPU)와 그래픽처리장치(GPU)가 함께 포함된 디바이스다. 홀로렌즈 앱을 개발하기 위해서는 Visual Studio 2017과 Unity 5.4 이상 버전과 홀로렌즈 포팅을 위한 윈도우 10 OS 환경이 필요하다.



<그림 2-15> 홀로렌즈[25]

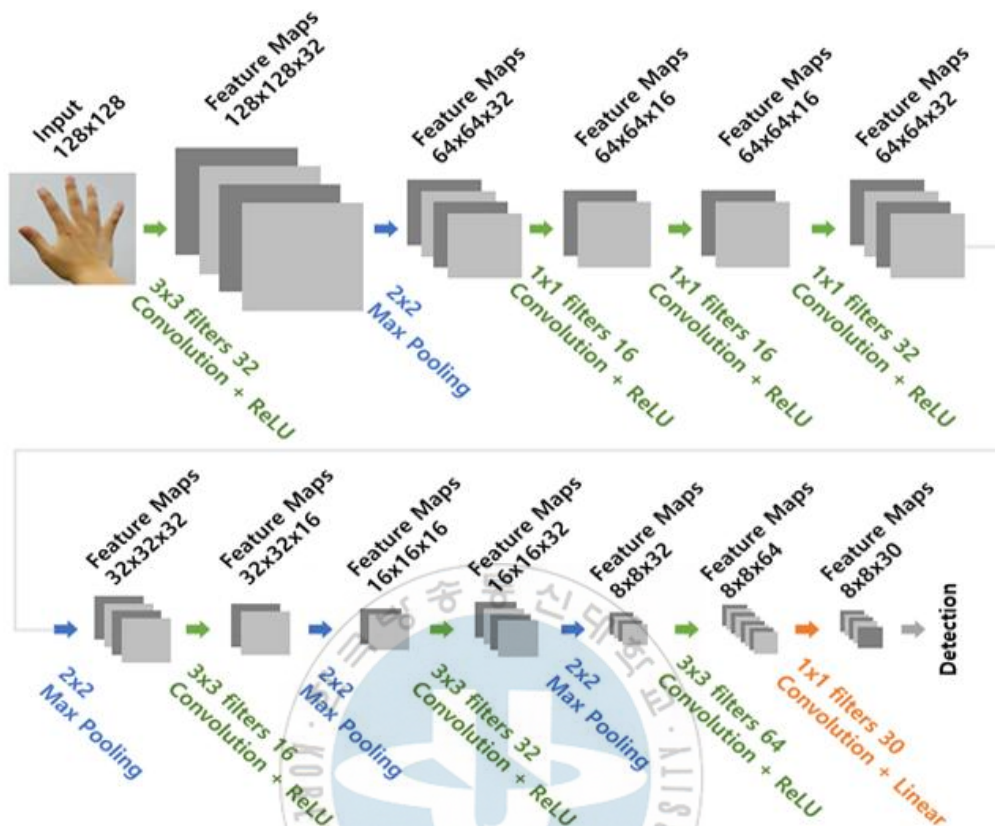
홀로렌즈 하드웨어 사양은 <표 2-3>과 같으며, 윈도우 10, 32-bit 운영체제를 사용하고 있어 딥러닝 프레임워크를 윈도우 DLL 컴파일 시 32-bit로 포팅해야 하는 등 개발 시 유의해야 한다.

<표 2-3> 홀로렌즈 하드웨어 사양[26]

항목	사양
OS	Windows 10.011802.1033 32-bit
CPU	Intel Atom x5-Z8100 1.04GHz IntelAirmont(14mm) 4LogicalProcessors64-bit capable
GPU/HPU	HoloLens Graphics
GPU Vender ID	8086h(Intel)
Dedicated Video Memory	114 MB
Shared System Memory	980 MB
RAM	2GB
Storage	64GB (54.09 GB available)
App Memory Usage Limit	900 MB
Camera Photos	2.4 MP (2048×1152)
Camera Video	1.1 MP (1408×792)
Video Speed	30 FPS

## (2) HoloLensHandNet

HoloLensHandNet 모델[27]은 홀로렌즈에서 실시간으로 구동 가능한 손 검출을 위한 딥러닝 모델이다. HoloLensHandNet 모델은 실시간 객체 검출 모델인 YOLO 모델 기반으로 제작되었으며, 홀로렌즈에 Darknet을 포팅하여 딥러닝 모델을 구동하였다. HoloLensHandNet 모델은 <그림 2-16>과 같은 아키텍처를 가지며, 홀로렌즈와 같은 웨어러블 저사양 디바이스에서 실시간 구동을 목적으로 설계되었다.



<그림 2-16> HoloLensHandNet 모델 아키텍처[27]

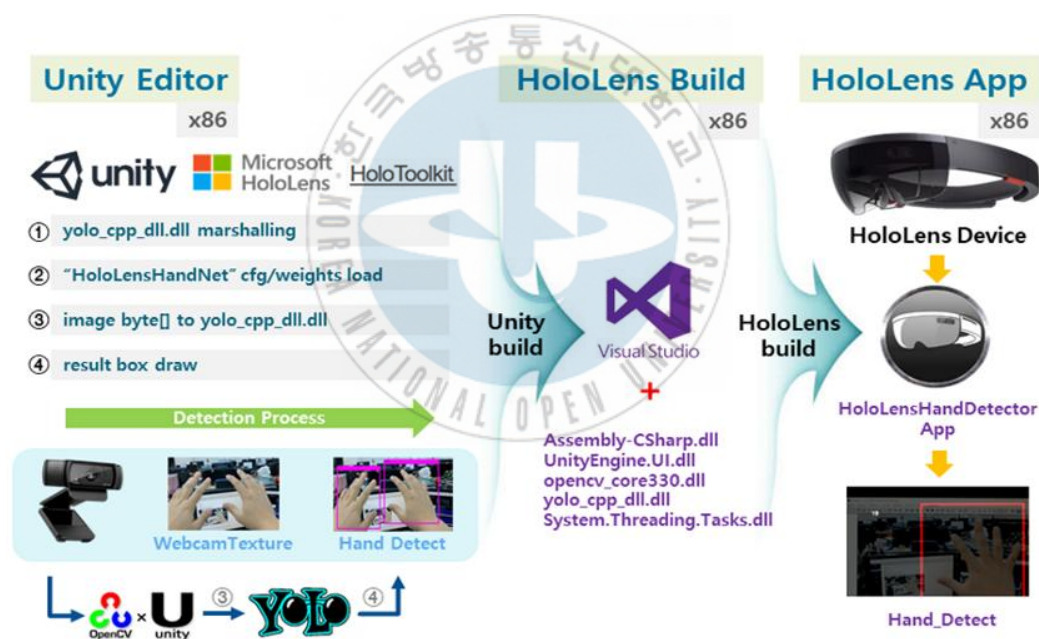
HoloLensHandNet 모델은 YOLO 모델 중에서 가장 경량화된 모델인 tiny-yolo 모델을 기반으로 연산량을 줄이는 방법으로 설계되었다. tiny-yolo 모델의 연산량은 5.41 BFLOPS이다. 실험 결과 연산량은 입력 이미지 사이즈와 모델 네트워크 파라미터에 따라 많은 영향을 받는다. HoloLensHandNet 모델의 입력 이미지 사이즈는 tiny-yolo 모델에서 제안된 416×416을 128×128로 변경하였지만, 정확도는 감소하지 않았다. 연산량을 개선하기 위해 모델 네트워크 구성에서 완전연결층을 제거하고, 컨볼루션 필터의 수를 줄이는 방법으로 개선하였다. 마지막으로 연산량을 개선하기 위해 딥러닝 네트워크 초기 위치한 3×3 컨볼루션 필터 대신 1×1 컨볼루션 필터를 배치하였다. HoloLensHandNet 모델의 활성화함수는 성능 향상을 위해 ReLU를 사용하였다.

손 데이터셋을 제작하여 훈련 및 평가 결과 정확도는 F1-score 0.83, mAP(mean



Average Precision) 88%, 평균 IoU 57.03%로 평가되었다. HoloLensHandNet 입력 사이즈 128 모델의 연산량은 0.054 BFLOPS로 tiny-yolo 입력 사이즈 416 모델과 비교하면, 연산량은 128.5배 적게 개선되어 CPU를 사용한 검출 시간은 10ms가 소요되었다. mAP는 16% 향상, F1-score는 5% 향상되었고, HoloLensHandNet 모델의 크기는 103kb로 경량화되었다.

홀로렌즈에서 구현된 손 검출기 앱의 아키텍처는 <그림 2-17>과 같다. 홀로렌즈 손 검출 앱 성능은 CPU 멀티스레드 기법을 적용하면 손 검출과 영역 표시까지 68ms 소요되고, 19 FPS로 처리할 수 있어 실시간 처리가 가능한 것으로 평가되었다. 이러한 성능은 증강/혼합현실용 HMD 웨어러블 디바이스에서 실시간 한국수화언어 통역 시스템 구동의 가능성을 보여준다.



<그림 2-17> 홀로렌즈 손 검출기 앱 아키텍처[27]

## 제3장 제안 방법

### 3.1. 한국수화언어 통역 시스템 개요

#### (1) 한국수화언어 통역 시스템 개요 및 구성안

한국수화언어 통역 시스템은 화자의 자연수화언어 동작(단어)을 통역하는 시스템으로 공간과 장소, 장비의 제약에서 자유로운 시스템을 제안하고자 한다.

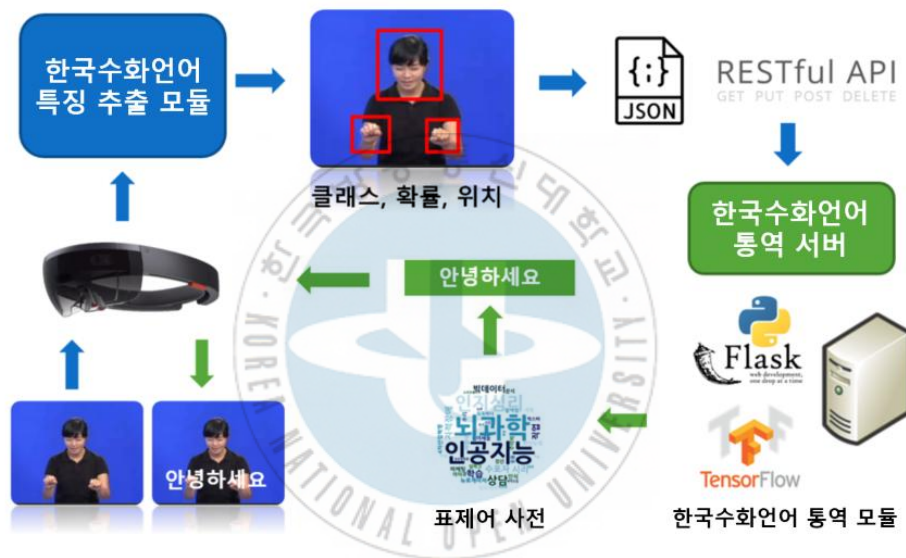
한국수화언어 통역 시스템은 한국수화언어 특징 추출 모듈과 한국수화언어 통역 모듈로 구성된다. 한국수화언어 특징 추출 모듈은 객체 검출 모듈로 손과 손가락 모양의 종류와 얼굴의 종류, 위치 정보를 추출할 수 있는 모듈이다. 한국수화언어 특징 추출 모듈은 증강/혼합현실용 HMD와 같은 웨어러블 플랫폼에 실시간 구동을 위해서는 빠른 성능과 경량화된 모듈 필요하다. 한국수화언어 통역 모듈은 한국수화언어 특징 추출 모듈에서 추출된 정보를 바탕으로 자연수화언어 동작(단어)을 통역하는 모듈이다. 한국수화언어의 특징 추출 모듈은 딥러닝의 객체 검출 모듈인 YOLO 모델을 기반으로 증강/혼합현실용 HMD와 같은 웨어러블 디바이스에서 빠르게 동작할 수 있도록 설계한다. 한국수화언어 통역 모듈은 CNN 알고리즘 중 DenseNet 모델을 적용하여 자연수화언어 동작(단어)을 표제어로 통역 모듈을 제안한다.

한국수화언어 통역 시스템은 증강/혼합현실용 HMD와 같은 웨어러블 디바이스에서 영상을 입력받아 실시간으로 특징을 추출할 수 있다. 추출된 한국수화언어 특징 정보를 한국수화언어 통역 모듈로 전송하는 것은 영상을 전송하는 것에 비해 통신 비용 절감의 효과가 있으며, 영상 데이터 전처리 없이 한국수화언어 특징 추출이 가능한 장점이 있다. 한국수화언어 통역 모듈은 한국수화언어 특징 정보를 한국수화언어 표제어로 통역한다.

한국수화언어 통역 시스템 구성안은 <그림 3-1>과 같은 시스템 아키텍처로 구성된다. 한국수화언어 통역 시스템을 데이터 흐름으로 보면 다음과 같다. 증강/혼합현실



용 HMD 웨어러블 디바이스 또는 PC의 웹캠에서 입력 영상 받아 임베딩된 한국수화언어 특징 추출 모듈에서 한국수화언어 특징 데이터를 손과 손가락 모양인 수형 종류, 얼굴 종류의 위치 및 확률 정보를 예측한다. 예측된 데이터는 JSON 포맷으로 가공하여 한국수화언어 통역 서버로 전달되고, 전처리 과정을 거쳐 112×112 사이즈의 특징맵으로 변환한다. 변환된 데이터는 한국수화언어 통역 모듈에 입력되어 한국수화언어 표제어와 확률값으로 출력되고, 표제어를 PC 혹은 웨어러블 디바이스 출력 장치에 전달하여 디스플레이한다.



<그림 3-1> 한국수화언어 통역 시스템 구성안

## (2) 한국수화언어 통역 시스템 개발환경









개발 환경은 증강/혼합현실용 HMD 웨어러블 디바이스인 홀로렌즈 포팅이 가능한 윈도우 10 OS 환경에 RTX-2080TI가 탑재된 PC와 딥러닝 모듈 훈련용 우분투 16.04 OS 환경 GTX-1080TI 4개가 장착된 서버를 사용하였다.

## 3.2. 한국수화언어 데이터셋




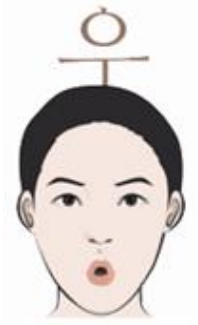
### (1) 한국수화언어 특징 추출 모듈 데이터셋

한국수화언어 특징 추출 모듈 데이터셋은 손과 손가락 모양을 기반으로 한 수형 65개와 손 모양이 아닌 비수지형 얼굴코드 6개로 구성되었다. 수형코드의 수형은 “수형 기반 한국수어사전 구축”의 수형코드[6]를 기반으로 제작되었다. 얼굴코드는 한국수화언어는 동작(단어)과 함께 단어 발음 입모양과 얼굴 표정을 함께 표현하기 때문에 한국수화언어 특징 추출 모듈에서는 이 정보를 포함해야 한다. 비수지형인 얼굴코드는 6개로 구성되었고, 기본 웃는 입 모양과 ‘아’, ‘에’, ‘이’, ‘오’, ‘우’ 발음의 얼굴 모양으로 얼굴코드로 구분하였다.

<표 3-1> 한국수화언어 수형코드 체계

ID	Code	수형 사진	이미지	분류	설명
0000	0			엄지와 손가락 손 끝을 붙인 수형	엄지와 네 손가락 손 끝을 붙인 수형
0001	1			엄지와 손가락 손 끝을 붙인 수형	엄지와 첫 번째 관절 을 구부려 손끝을 붙 인 수형
0002	2			엄지와 손가락 손 끝을 붙인 수형	엄지와 검지 손끝을 붙이고 중지와 약지 와 소지를 접은 수형
...					
1200	64			엄지와 소지를 가 진 수형	엄지와 소지를 펴고 검지와 중지와 약지 를 접어 손바닥에 손 끝을 접촉한 수형

<표 3-2> 한국수화언어 얼굴코드 체계

ID	Code	이미지	분류	설명
2000	65		얼굴인식	기본 얼굴
2001	66		얼굴인식	“아” 발음 얼굴
2002	67		얼굴인식	“에” 발음 얼굴
...				
2005	70		얼굴인식	“우” 발음 얼굴

한국수화언어 특징 추출 데이터셋은 <그림 3-2>와 같이 Train(6,917개), Test(768개)가 제작되었으며, 한국수화언어 특징 추출 데이터셋 제작 프로세스는 다음과 같다.

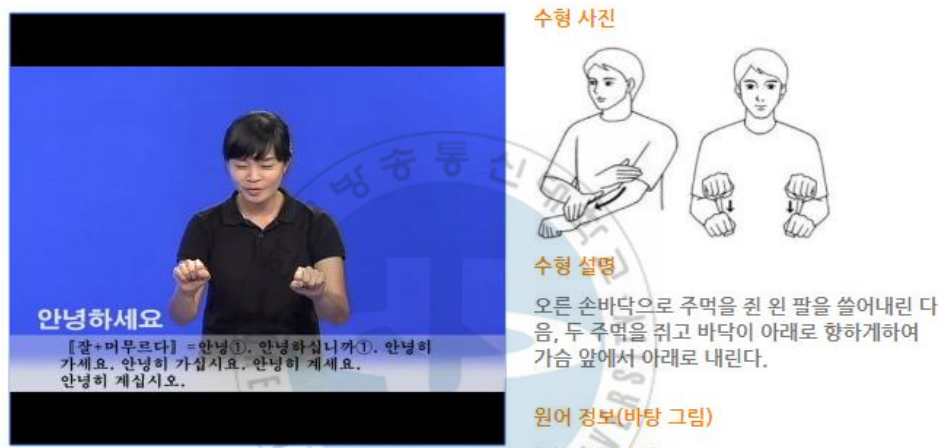
- ① 각 수형코드 및 얼굴코드 동영상 촬영
- ② 각 코드별 동영상 편집
- ③ 각 코드별 동영상 이미지로 변환(각 코드별 약 500장)
- ④ 각 코드별 이미지를 5등분으로 분리하여 데이터셋 버전 01번 제작(7,685개)
- ⑤ 박싱 라벨링 툴을 사용하여 태깅(7,685개)[28]
- ⑥ 좌표 태깅된 텍스트 파일을 darknet 포맷(class, x1, y1, x2, y2)으로 변경
- ⑦ Train(6,917개), Test(768개) 데이터 분리
- ⑧ 제작된 이미지셋과 태깅된 텍스트 파일 적용



<그림 3-2> 한국수화언어 특징 추출 데이터셋

## (2) 한국수화언어 통역 모듈 데이터셋

한국수화언어 통역 모듈 데이터셋은 한국수어사전에서 일상생활 수화언어 16개의 카테고리로 1,210개 표제어와 <그림 3-3>과 같은 동영상상을 수집하였다[29]. <표 3-3>과 같이 한국수화언어 일상생활 대표 표제어 10개를 선정하였고, <표 3-4>와 같이 데이터셋 수집 리스트를 작성하였다. 마지막으로 <그림 3-4> 한국수화언어 일상생활 대표 표제어 데이터셋 120개를 제작하였다.






<그림 3-3> 한국수어사전 수화언어 영상[29]






한국수화언어 일상생활 대표 표제어 데이터셋 수집 환경은 다음과 같다.

- ① 다운로드 영상 일반 속도 수화언어[29]
- ② 다운로드 영상 느린 속도 수화언어[29]
- ③ 캠코더 환경 그린 배경 정면 및 측면 수화언어
- ④ 캠코더 환경 화이트 배경 정면 및 측면 수화언어
- ⑤ 캠코더 환경 복잡한 배경 정면 및 측면 수화언어
- ⑥ 웹캠 환경 하이앵글 배경 정면 및 측면 수화언어
- ⑦ 웹캠 환경 로우앵글 배경 정면 및 측면 수화언어

<표 3-3> 한국수화언어 일상생활 대표 표제어

No	ID	vocab ID	표제어	표제어 설명	수화언어 이미지	수화언어 설명
1	1000	0	고맙다	[형용사] 남이 베풀어 준 호의나 도움 따위에 대하여 마음이 흐뭇하고 즐겁다.		손끝이 밖으로 향하게 펴서 모로 세운 오른손으로 손바닥이 아래로 향하게 편 왼손등을 두 번 두드린다.
2	1001	1	괜찮다	[형용사] 별로 나쁘지 않고 보통 이상이다.		오른 주먹의 4지를 펴서 끝 바닥을 턱에 가볍게 두 번 댄다.
3	1002	2	사랑	[명사] 남녀 간에 그리워하거나 좋아하는 마음. 또는 그런 일.		5지 바닥을 1지 옆면에 대고 손등이 왼쪽으로 향하게 모로 세운 왼 주먹 위에 오른 손바닥을 대고 오른손만 오른쪽으로 돌린다.
4	1003	3	기쁘다	[형용사] 욕구가 충족되어 마음이 흐뭇하고 흡족하다.		두 주먹의 1·5지를 펴고 손등이 밖으로 향하게 하여 1지 끝을 양쪽 가슴 앞에 대고 상하로 두 번 엇갈리게 움직인다.
5	1004	4	미안	[명사] 남에게 대하여 마음이 편치 못하고 부끄러움.		오른손의 1·5지 끝을 맞대어 동그라미를 만들어 이마에 댔다가 1·5지를 펴며 내린다.

(표 계속)

6	1005	5	안되다	[동사] 일, 현상, 물건 따위가 좋게 이루어지지 않다.		오른손의 2·5지 끝을 맞대어 동그라미를 만들고 1지를 코 옆에 댔다가 밖으로 내리며 2지를 5지에 뿔겨 편다.
7	1006	6	안녕	[감탄사] 편한 사이에서, 서로 만나거나 헤어질 때 정답게 하는 인사말.		"두 주먹을 쥐고 바닥이 아래로 향하게 하여 가슴 앞에서 아래로 내린다.
8	1007	7	늦다	[형용사] 기준이 되는 때보다 뒤져 있다.		손등이 위로 향하게 편 왼 손등에 오른손을 모로 세워 천천히 밖으로 내민다.
9	1008	8	만나다	[동사] 누군가가거나 와서 둘이 서로 마주 보다.		두 주먹의 1지를 펴서 마주 세웠다가 중앙으로 모아 마주 댄다.
10	1009	9	부끄럽다	[형용사] 스스럼을 느끼어 매우 수줍다.		오른손의 1·5지 끝을 맞대어 동그라미를 만들어 세워 등을 오른쪽 볼에 두 번 댔다 뼀다.



<표 3-4> 한국수화언어 일상생활 대표 표제어 데이터셋 리스트

No	ID	vocab ID	표제어	동영상	표제어 설명	촬영 환경
1	1000	0	고맙다	1000.mp4	[형용사] 남이 베풀어 준 호의나 도움 따위에 대하여 마음이 흐뭇하고 즐겁다.	다운로드
2	1001	1	괜찮다	1001.mp4	[형용사] 별로 나쁘지 않고 보통 이상이다.	다운로드
3	1002	2	사랑	1002.mp4	[명사] 남녀 간에 그리워하거나 좋아하는 마음. 또는 그런 일.	다운로드
4	1003	3	기쁘다	1003.mp4	[형용사] 욕구가 충족되어 마음이 흐뭇하고 흡족하다.	다운로드
5	1004	4	미안	1004.mp4	[명사] 남에게 대하여 마음이 편치 못하고 부끄러움.	다운로드
6	1005	5	안되다	1005.mp4	[동사] 일, 현상, 물건 따위가 좋게 이루어지지 않다.	다운로드
7	1006	6	안녕	1006.mp4	[감탄사] 편한 사이에서, 서로 만나거나 헤어질 때 정답게 하는 인사말.	다운로드
8	1007	7	늦다	1007.mp4	[형용사] 기준이 되는 때보다 뒤져 있다.	다운로드
9	1008	8	만나다	1008.mp4	[동사] 누군가 가거나 와서 둘이 서로 마주 보다.	다운로드
10	1009	9	부끄럽다	1009.mp4	[형용사] 스스러움을 느끼어 매우 수줍다.	다운로드
...						
111	1510	0	고맙다	1510.mp4	[형용사] 남이 베풀어 준 호의나 도움 따위에 대하여 마음이 흐뭇하고 즐겁다.	웹캠/로우앵글
112	1511	1	괜찮다	1511.mp4	[형용사] 별로 나쁘지 않고 보통 이상이다.	웹캠/로우앵글
113	1512	2	사랑	1512.mp4	[명사] 남녀 간에 그리워하거나 좋아하는 마음. 또는 그런 일.	웹캠/로우앵글
114	1513	3	기쁘다	1513.mp4	[형용사] 욕구가 충족되어 마음이 흐뭇하고 흡족하다.	웹캠/로우앵글
115	1514	4	미안	1514.mp4	[명사] 남에게 대하여 마음이 편치 못하고 부끄러움.	웹캠/로우앵글
116	1515	5	안되다	1515.mp4	[동사] 일, 현상, 물건 따위가 좋게 이루어지지 않다.	웹캠/로우앵글
117	1516	6	안녕	1516.mp4	[감탄사] 편한 사이에서, 서로 만나거나 헤어질 때 정답게 하는 인사말.	웹캠/로우앵글
118	1517	7	늦다	1517.mp4	[형용사] 기준이 되는 때보다 뒤져 있다.	웹캠/로우앵글
119	1518	8	만나다	1518.mp4	[동사] 누군가 가거나 와서 둘이 서로 마주 보다.	웹캠/로우앵글
120	1519	9	부끄럽다	1519.mp4	[형용사] 스스러움을 느끼어 매우 수줍다.	웹캠/로우앵글



<그림 3-4> 한국수화언어 일상생활 대표 표제어 데이터셋

### 3.3. 한국수화언어 특징 추출 모듈

#### (1) 한국수화언어 특징 추출 모듈 설계 및 구현

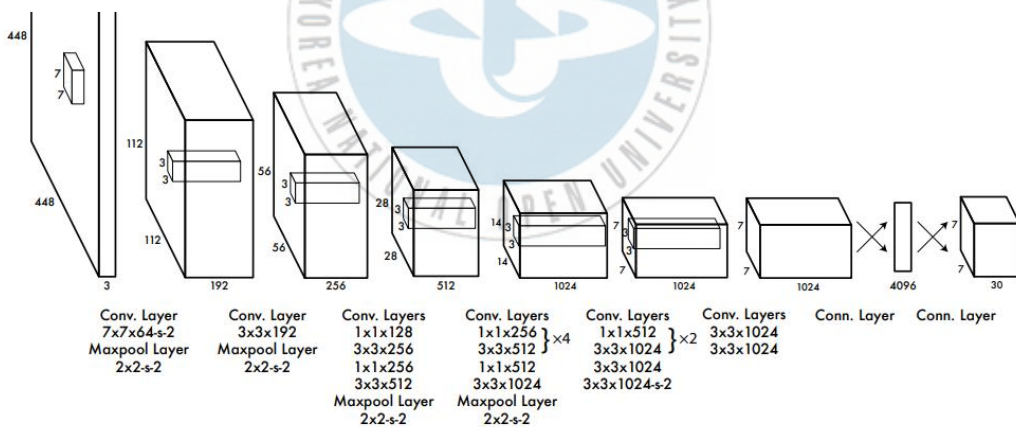
한국수화언어 특징 추출 모듈은 딥러닝을 이용한 실시간 객체 검출 모델인 YOLO 모델 기반으로 제작되었으며, YOLO 모델 아키텍처는 <그림 3-5>와 같다.

KSL-yolov3-tiny 특징 추출 모듈은 YOLO 모델 중 정확도가 높은 yolov3-tiny 모델 기반으로 한국수화언어 특징 추출을 위한 객체 검출을 목적으로 딥러닝 네트워크 구성 및 파라미터를 설정하였고, 세부 파라미터는 <표 3-5>로 정리하였다. yolov3-tiny 모듈은 연산량 때문에 증강/혼합현실용 HMD와 같은 웨어러블 디바이스에서는 실시간 객체 검출에는 어려움이 있지만, 한국수화언어 통역 모듈을 위한 한국수화언어

특징 데이터셋 수집에 활용할 수 있다.

KSL-yolov3-tiny 특징 추출 모듈의 성능 개선을 위해 한국수화언어 특징 추출 모듈인 KSL-HoloLensNet을 제안한다. KSL-HoloLensNet 모듈은 증강/혼합현실용 HMD 디바이스인 홀로렌즈에서 19 FPS의 성능을 갖는 HoloLensHandNet 모델을 바탕으로 제작되었다. 제안된 KSL-HoloLensNet 모듈은 영상에서 손을 검출하는 HoloLensHandNet과 달리 한국수화언어 특징인 수형코드 및 얼굴코드의 71개 클래스를 예측할 수 있다. KSL-HoloLensNet 모듈은 입력 이미지 사이즈를 352로 제안하였다. 이는 입력 이미지 사이즈가 352 이하로 작아지면 정확도가 감소했기 때문이다. KSL-HoloLensNet은 연산량 BFLOPS를 줄이기 위해 딥러닝 네트워크 초기에 1×1 컨볼루션 필터 배치하여 딥러닝 네트워크에서 컨볼루션 레이어 증가에 따른 파라미터 증가를 최소화하였다. KSL-HoloLensNet 모듈의 세부 파라미터는 표 <표 3-6>으로 정리하였다.

한국수화언어 특징 추출 모듈을 사용하면 <그림 3-6>과 같이 한국수화언어 특징의 클래스 확률 정보 및 위치 정보를 예측하고 추출할 수 있다.



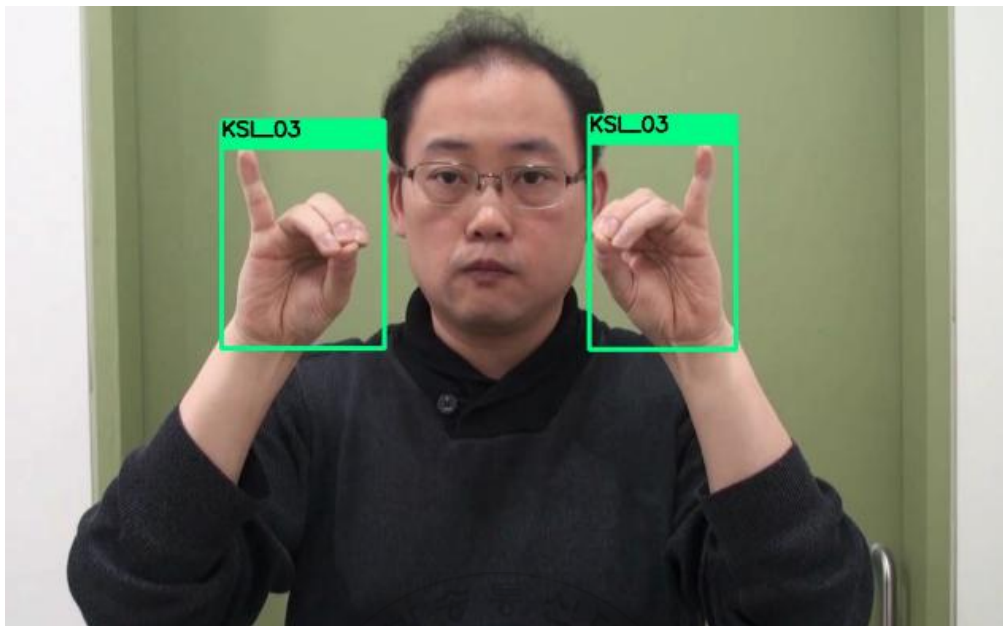
<그림 3-5> YOLO 아키텍처[20]

<표 3-5> KSL-yolov3-tiny 모듈 파라미터

No	Layer	Filters	Size	Input	Output	BFLOPS
0	conv	16	3 x 3 / 1	416 x 416 x 3	416 x 416 x 16	0.150
1	max		2 x 2 / 2	416 x 416 x 16	208 x 208 x 16	0.003
2	conv	32	3 x 3 / 1	208 x 208 x 16	208 x 208 x 32	0.399
3	max		2 x 2 / 2	208 x 208 x 32	104 x 104 x 32	0.001
4	conv	64	3 x 3 / 1	104 x 104 x 32	104 x 104 x 64	0.399
5	max		2 x 2 / 2	104 x 104 x 64	52 x 52 x 64	0.001
6	conv	128	3 x 3 / 1	52 x 52 x 64	52 x 52 x 128	0.399
7	max		2 x 2 / 2	52 x 52 x 128	26 x 26 x 128	0.000
8	conv	256	3 x 3 / 1	26 x 26 x 128	26 x 26 x 256	0.399
9	max		2 x 2 / 2	26 x 26 x 256	13 x 13 x 256	0.000
10	conv	512	3 x 3 / 1	13 x 13 x 256	13 x 13 x 512	0.399
11	max		2 x 2 / 1	13 x 13 x 512	13 x 13 x 512	0.000
12	conv	1024	3 x 3 / 1	13 x 13 x 512	13 x 13 x 1024	1.595
13	conv	256	1 x 1 / 1	13 x 13 x 1024	13 x 13 x 256	0.089
14	conv	512	3 x 3 / 1	13 x 13 x 256	13 x 13 x 512	0.399
15	conv	228	1 x 1 / 1	13 x 13 x 512	13 x 13 x 228	0.039
16	yolo					
17	route	13				
18	conv	128	1 x 1 / 1	13 x 13 x 256	13 x 13 x 128	0.011
19	upsample		2x	13 x 13 x 128	26 x 26 x 128	
20	route	19 8				
21	conv	256	3 x 3 / 1	26 x 26 x 384	26 x 26 x 256	1.196
22	conv	228	1 x 1 / 1	26 x 26 x 256	26 x 26 x 228	0.079
23	yolo					

<표 3-6> KSL-HoloLensNet 모듈 파라미터

No	Layer	Filters	Size	Input	Output	BFLOPS
0	conv	8	3 x 3 / 1	352 x 352 x 3	352 x 352 x 8	0.054
1	conv	16	1 x 1 / 1	352 x 352 x 8	352 x 352 x 16	0.032
2	max		2 x 2 / 2	352 x 352 x 16	176 x 176 x 16	0.002
3	conv	16	3 x 3 / 1	176 x 176 x 16	176 x 176 x 16	0.143
4	max		2 x 2 / 2	176 x 176 x 16	88 x 88 x 16	0.000
5	conv	32	3 x 3 / 1	88 x 88 x 16	88 x 88 x 32	0.071
6	max		2 x 2 / 2	88 x 88 x 32	44 x 44 x 32	0.000
7	conv	64	3 x 3 / 1	44 x 44 x 32	44 x 44 x 64	0.071
8	max		2 x 2 / 2	44 x 44 x 64	22 x 22 x 64	0.000
9	conv	128	3 x 3 / 1	22 x 22 x 64	22 x 22 x 128	0.071
10	max		2 x 2 / 2	22 x 22 x 128	11 x 11 x 128	0.000
11	conv	256	3 x 3 / 1	11 x 11 x 128	11 x 11 x 256	0.071
12	max		2 x 2 / 2	11 x 11 x 256	6 x 6 x 256	0.000
13	conv	512	3 x 3 / 1	6 x 6 x 256	6 x 6 x 512	0.085
14	conv	380	1 x 1 / 1	6 x 6 x 512	6 x 6 x 380	0.014
15	detection					



<그림 3-6> 한국수화언어 특징 추출 모듈 실행 화면

## (2) 한국수화언어 특징 데이터 추출

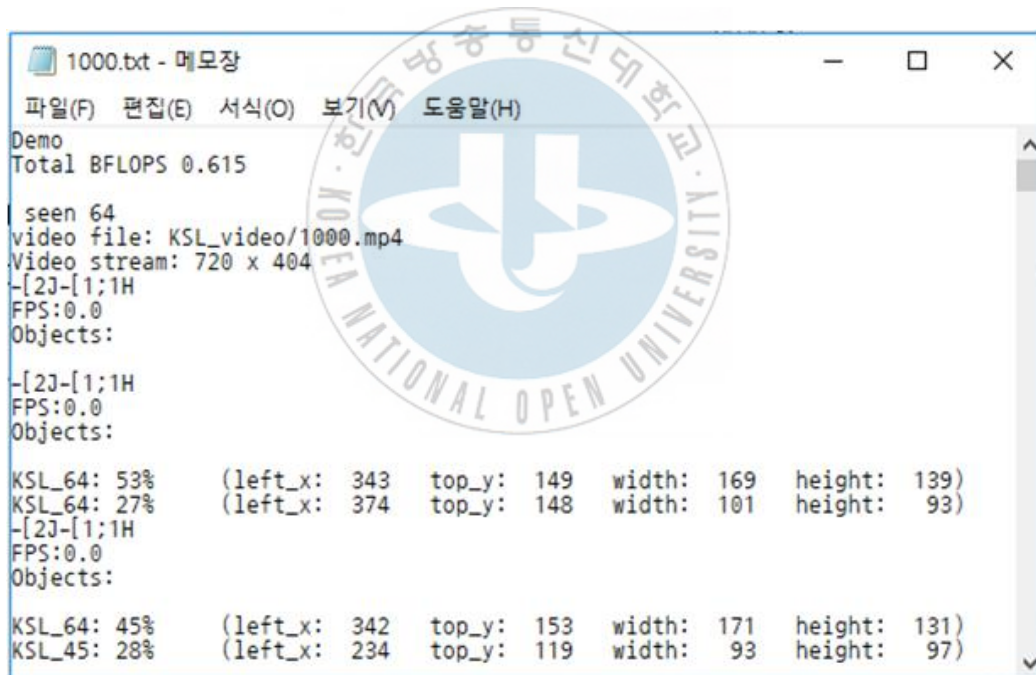
한국수화언어 특징 추출 모듈을 사용하여 한국수화언어 특징 데이터를 추출할 수 있다. 한국수화언어 특징 추출 모듈을 사용한 한국수화언어 특징 데이터 추출은 오픈소스 신경망 프레임워크인 Darknet[22]이 사용된다.

KSL-yolov3-tiny 한국수화언어 특징 추출 모듈과 KSL-HoloLensNet 한국수화언어 특징 추출 모듈을 Darknet 명령어를 사용하여 120개의 한국수화언어 대표 표제어 데이터셋 영상에서 한국수화언어 특징 데이터를 추출한다. KSL-yolov3-tiny 한국수화언어 특징 추출 모듈과 KSL-HoloLensNet 한국수화언어 특징 추출 모듈의 threshold 값을 설정하여 다양한 한국수화언어 특징 데이터를 추출할 수 있으며, 추출된 데이터는 데이터 프로세싱 과정 후 한국수화언어 통역 모델의 훈련 및 평가 데이터셋으로 사용된다. KSL-HoloLensNet 한국수화언어 특징 추출 모듈로 한국수화언어 특징을 추출하는 Darknet 명령어는 다음과 같다.



```
darknet.exe      detector      demo      cfg/KSL.data      cfg/KSL-HoloLensNet.cfg
KSL-HoloLensNet.weights  KSL_video/1000.mp4  -i  0  -thresh  0.25  -out_filename
KSL_video/1000.mp4 -dont_show -ext_output <KSL_video/1000.txt> KSL_video/1000.txt
```

한국수화언어 대표 표제어 120개의 영상을 대상으로 KSL-HoloLensNet 한국수화언어 특징 추출 모듈을 Darknet으로 구동하여 특징을 추출하면 120개의 텍스트 파일로 추출된다. 추출된 데이터는 <그림 3-7>과 같고, 추출 정보는 video file, Video stream 크기, Objects 하위의 검출된 한국수화언어 특징 데이터로 구성된다. 한국수화언어 특징 데이터는 ‘KSL\_46:36%’와 같은 한국수화언어 특징 ID와 확률 정보 및 ‘(left\_x: 343 top\_y: 149 width: 169 height: 139)’와 같은 위치, 크기를 포함한다.



<그림 3-7> 한국수화언어 특징 추출 데이터

### 3.4. 한국수화언어 통역 모듈

#### (1) 한국수화언어 통역 모듈 설계 및 구현

한국수화언어 통역 모듈은 CNN 알고리즘 기반 DesneNet[24]을 사용하였다. CNN은 이미지의 공간 정보를 유지하면서 인접 이미지와의 특징을 효과적으로 인식하고 강조하는 방식으로 이미지의 특징을 추출하는 부분과 이미지를 분류하는 부분으로 구성된다. 특징 추출 영역은 필터를 사용하여 공유 파라미터 수를 최소화하면서 이미지의 특징을 찾는 컨볼루션 레이어와 특징을 강화하고 모으는 풀링 레이어로 구성된다. CNN은 필터의 크기, 스프라이드, 패딩과 풀링 크기로 출력 데이터 크기를 조절하고, 필터의 개수로 출력 데이터의 채널을 결정한다. CNN 알고리즘은 네트워크가 깊어질수록 처음에 가지고 있던 정보가 사라지는 “wash out”의 문제점을 갖고 있다. 이 문제점과 성능을 개선을 위해 DesneNet이 제안되었다.

본 논문의 한국수화언어 통역 모듈은 딥러닝 기법 중 DesneNet 구현 소스[30]를 참고하여 구현하였다. 한국수화언어 통역 모듈의 훈련 데이터는 KSL-yolov3-tiny 한국수화언어 특징 추출 모듈에서 추출된 한국수화언어 특징 데이터와 KSL-HoloLensNet 한국수화언어 특징 추출 모듈에서 추출된 한국수화언어 특징 데이터를 사용하여 만들었다.

한국수화언어 통역 모듈의 구현 과정은 ① 훈련 및 평가 데이터 가공, ② 데이터 전처리, ③ DesneNet 적용, ④ 훈련, ⑤ 검증 모듈 제작 과정으로 구현되었다.

① 훈련 및 평가 데이터 가공 과정은 한국수화언어 통역 모듈 데이터셋 120개를 한국수화언어 특징 추출 모듈을 사용하여 한국수화언어 특징 데이터로 추출하고, 이를 JSON 데이터로 통합한다. KSL-yolov3-tiny 한국수화언어 특징 추출 모듈로 추출된 한국수화언어 특징 데이터 112개 아이템을 KSL\_NLP\_dataset\_yolov3.json 파일로 통합 변환하였다.

한국수화언어 특징 데이터로 통합된 JSON 데이터는 <표 3-7>와 같이 한국수화언어 일상생활 대표 표제어 ID를 나타내는 ‘ksl\_nlp\_id’, 파일명 ‘txt’, 한국수화언어 추



출 영상명 'video', 영상의 폭 'video\_w', 영상의 높이 'video\_h', 영상의 프레임 수 'frame'의 정보와 한국수화언어 특징 데이터를 담는 'data'로 구성되어 있다. 한국수화언어 특징을 담는 'data'의 데이터 구조는 'KSL\_id', 'prob', 'left\_x', 'top\_y', 'width', 'height'의 한국수화언어 특징 ID와 확률, 위치 및 크기 정보를 포함하는 구조를 갖는다.

<표 3-7> KSL\_NLP\_dataset\_yolov3.json 예시

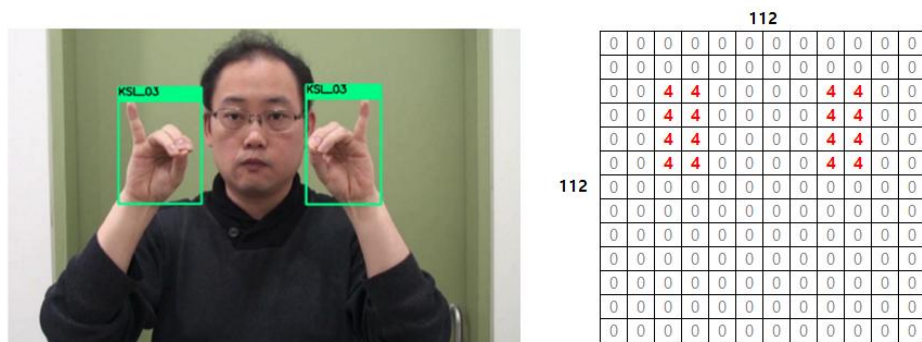
No	한국수화언어 특징 JSON 데이터
1	{'ksl_nlp_id': 1, 'txt': '1001.txt', 'video': '1001.mp4', 'video_w': 720, 'video_h': 404, 'frame': 42, 'data': [[], [], [], [], [], [{'KSL_id': 'KSL_70', 'prob': 10, 'left_x': 300, 'top_y': 52, 'width': 93, 'height': 132}], [{'KSL_id': 'KSL_70', 'prob': 16, 'left_x': 300, 'top_y': 48, 'width': 91, 'height': 136}], {'KSL_id': 'KSL_50', 'prob': 12, 'left_x': 298, 'top_y': 108, 'width': 63, 'height': 102}], [{'KSL_id': 'KSL_70', 'prob': 11, 'left_x': 300, 'top_y': 50, 'width': 91, 'height': 137}], {'KSL_id': 'KSL_50', 'prob': 13, 'left_x': 310, 'top_y': 92, 'width': 61, 'height': 114}], [{'KSL_id': 'KSL_50', 'prob': 16, 'left_x': 311, 'top_y': 92, 'width': 62, 'height': 113}], [], [], [], [], [], [], [], [], [], [{'KSL_id': 'KSL_70', 'prob': 10, 'left_x': 301, 'top_y': 56, 'width': 91, 'height': 124}], [], [], [], [], [], [], [], [], [], [{'KSL_id': 'KSL_70', 'prob': 14, 'left_x': 301, 'top_y': 50, 'width': 94, 'height': 129}], [], [], []]}
2	{'ksl_nlp_id': 2, 'txt': '1002.txt', 'video': '1002.mp4', 'video_w': 720, 'video_h': 404, 'frame': 54, 'data': [[], [], [{'KSL_id': 'KSL_70', 'prob': 54, 'left_x': 334, 'top_y': 10, 'width': 119, 'height': 158}], [{'KSL_id': 'KSL_70', 'prob': 46, 'left_x': 334, 'top_y': 10, 'width': 121, 'height': 158}], [{'KSL_id': 'KSL_70', 'prob': 41, 'left_x': 334, 'top_y': 10, 'width': 119, 'height': 159}], [{'KSL_id': 'KSL_70', 'prob': 44, 'left_x': 333, 'top_y': 10, 'width': 119, 'height': 159}], [{'KSL_id': 'KSL_70', 'prob': 38, 'left_x': 333, 'top_y': 10, 'width': 119, 'height': 159}], [{'KSL_id': 'KSL_70', 'prob': 39, 'left_x': 334, 'top_y': 10, 'width': 119, 'height': 159}], [{'KSL_id': 'KSL_70', 'prob': 42, 'left_x': 334, 'top_y': 10, 'width': 119, 'height': 159}], {'KSL_id': 'KSL_60', 'prob': 14, 'left_x': 339, 'top_y': 248, 'width': 83, 'height': 72}]}}

KSL\_NLP\_dataset.json 한국수화언어 특징 데이터셋은 KSL-yolov3-tiny 한국수화언어 특징 추출 모듈과 KSL-HoloLensNet 한국수화언어 특징 추출 모듈의 threshold 값을 다양한 예측값으로 조정하여 구성하였다. KSL\_NLP\_dataset.json 데이터셋에 사용된 threshold 값은 0.25, 0.30이며, 464개의 한국수화언어 특징 추출 데이터 아이템으로 통합 변환하였다. KSL\_NLP\_dataset.json 한국수화언어 특징 데이터셋은 <표 3-8>과 같이 KSL\_NLP\_dataset\_yolov3.json와 JSON 구조는 같고, 한국수화언어 특징 데이터는 한국수화언어 특징을 다양하게 인식하고 있다.

<표 3-8> KSL\_NLP\_dataset.json 예시

No	한국수화언어 특징 JSON 데이터
1	{'ksl_nlp_id': 0, 'txt': '1000.txt', 'video': '1000.mp4', 'video_w': 720, 'video_h': 404, 'frame': 49, 'data': [[], [], [{ 'KSL_id': 'KSL_64', 'prob': 53, 'left_x': 343, 'top_y': 149, 'width': 169, 'height': 139}, { 'KSL_id': 'KSL_64', 'prob': 27, 'left_x': 374, 'top_y': 148, 'width': 101, 'height': 93}], [{ 'KSL_id': 'KSL_64', 'prob': 45, 'left_x': 342, 'top_y': 153, 'width': 171, 'height': 131}, { 'KSL_id': 'KSL_45', 'prob': 28, 'left_x': 234, 'top_y': 119, 'width': 93, 'height': 97}], [{ 'KSL_id': 'KSL_64', 'prob': 39, 'left_x': 370, 'top_y': 169, 'width': 120, 'height': 99}, { 'KSL_id': 'KSL_64', 'prob': 34, 'left_x': 365, 'top_y': 148, 'width': 99, 'height': 89}], [{ 'KSL_id': 'KSL_64', 'prob': 45, 'left_x': 365, 'top_y': 148, 'width': 101, 'height': 87}, { 'KSL_id': 'KSL_64', 'prob': 36, 'left_x': 336, 'top_y': 150, 'width': 179, 'height': 130}], [{ 'KSL_id': 'KSL_64', 'prob': 50, 'left_x': 362, 'top_y': 149, 'width': 106, 'height': 87}, { 'KSL_id': 'KSL_64', 'prob': 36, 'left_x': 334, 'top_y': 152, 'width': 177, 'height': 127}, { 'KSL_id': 'KSL_52', 'prob': 32, 'left_x': 235, 'top_y': 119, 'width': 93, 'height': 96}], [{ 'KSL_id': 'KSL_64', 'prob': 54, 'left_x': 360, 'top_y': 150, 'width': 108, 'height': 86}, { 'KSL_id': 'KSL_64', 'prob': 34, 'left_x': 334, 'top_y': 154, 'width': 173, 'height': 125}], [{ 'KSL_id': 'KSL_64', 'prob': 59, 'left_x': 365, 'top_y': 150, 'width': 109, 'height': 88}, { 'KSL_id': 'KSL_64', 'prob': 29, 'left_x': 335, 'top_y': 155, 'width': 172, 'height': 124}, { 'KSL_id': 'KSL_50', 'prob': 25, 'left_x': 239, 'top_y': 127, 'width': 95, 'height': 98}], [{ 'KSL_id': 'KSL_64', 'prob': 64, 'left_x': 372, 'top_y': 154, 'width': 104, 'height': 82}, { 'KSL_id': 'KSL_64', 'prob': 31, 'left_x': 343, 'top_y': 155, 'width': 165, 'height': 124}]]}

② 데이터 전처리 과정은 한국수화언어 특징 데이터셋으로 변환된 JSON 데이터를 기반으로 DesneNet 훈련하기 위한 112×112 특징맵 데이터로 변환한다. <그림 3-8>과 같이 112×112 특징맵은 기본값을 0으로 설정한 후 가장 확률이 높은 값의 인덱스 1~71을 채우는 방법으로 구현되었다.



<그림 3-8> 한국수화언어 특징맵 예시

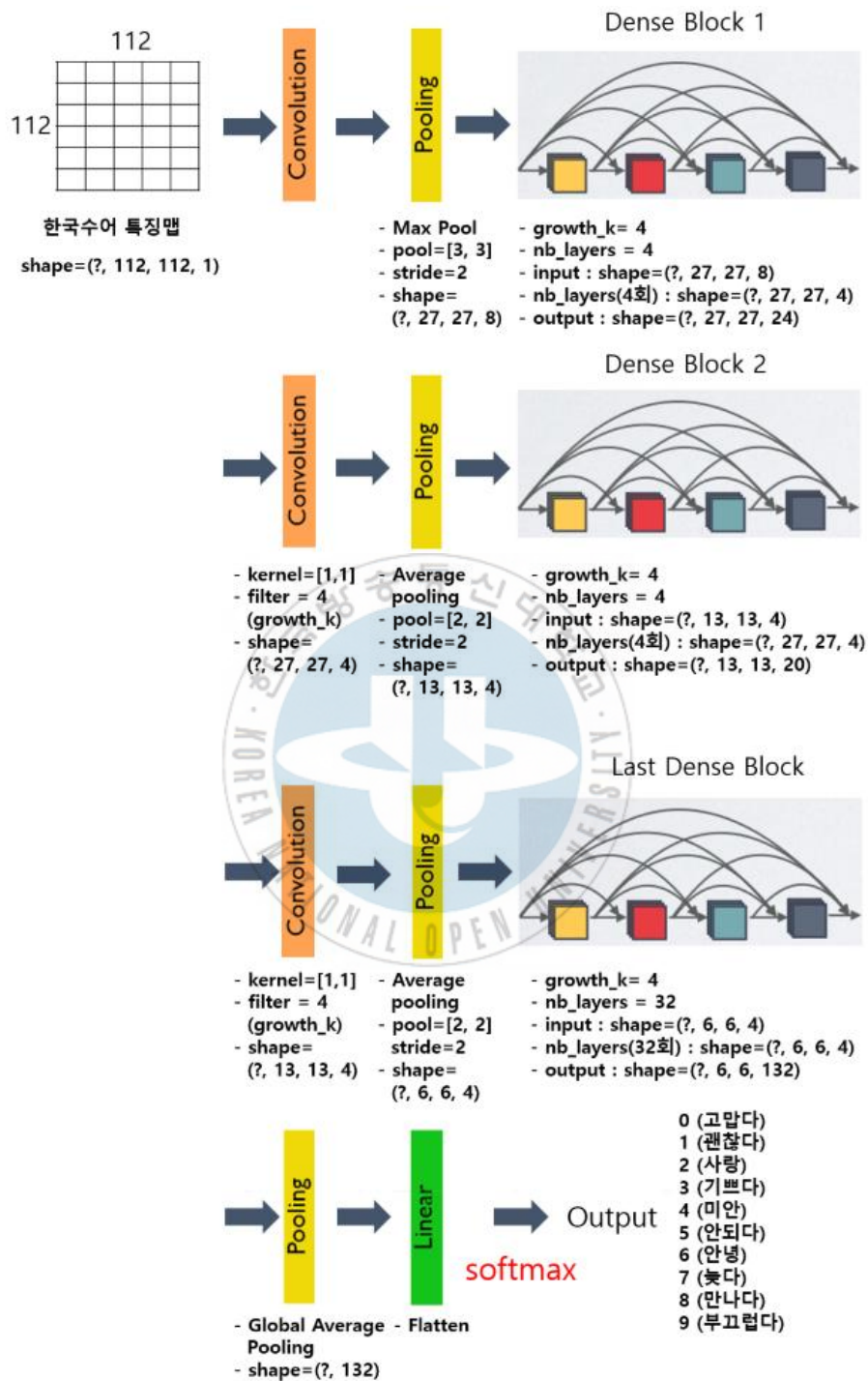
③ DesneNet 적용 과정은 다음과 같다. CNN은 네트워크가 깊어질수록 처음에 가지고 있던 정보가 사라지는 “wash out” 문제가 발생한다. DenseNet에서는 growth\_k(성장률)을 사용하여 처음 쌓은 층을 네트워크가 깊어져도 계속 차곡차곡 쌓아가는 것으로 이러한 문제를 해결하였다.

한국수화언어 통역 모듈의 DenseNet 모듈에 사용된 파라미터를 <표 3-9>에 정리하였다. DenseNet 모듈에 사용된 파라미터는 배치 크기, 데이터셋 반복 횟수, 한국수화언어 특징 차원수, 프레임 폭, 프레임 높이, 임베딩 차원 수, DenseNet 입력의 한 변 크기, 표제어 클래스 수, dense block 반복 수, DenseNet growth\_k(성장률)이다. dense block 반복 수를 늘리면 깊은 네트워크 구성이 가능하며, DenseNet growth\_k 조절을 통해 정보의 밀집도를 높일 수 있다.

한국수화언어 통역 모듈에 사용한 DenseNet 모듈 아키텍처는 <그림 3-9>와 같다. 한국수화언어 통역 모델에서는 전 처리된 112×112 사이즈의 한국수화언어 특징맵 입력값을 가지며, 출력값은 한국수화언어 일상생활 대표 표제어 인덱스 0~9의 값을 갖는다. 한국수화언어 통역 모듈은 3개의 Dense Block을 사용하며 growth\_k=4를 사용하여 네트워크 파라미터 연산 증가를 최소화하였다.

<표 3-9> 한국수화언어 통역 모듈 사용 파라미터

파라미터	값	설명
batch_size	512	배치 크기
num_epochs	1001	데이터셋 반복 횟수
ksl_dim	71	한국수화언어 특징 차원 수
frame_width	720	프레임 폭
frame_height	404	프레임 높이
embedding_size	1	임베딩 차원 수
input_size	112	DenseNet 입력의 한 변 크기
n_class	10	표제어 클래스 수
nb_block	2	dense block 반복 수
growth_k	4	DenseNet growth_k(성장률)



<그림 3-9> 한국수화언어 통역 DesNet 모듈 아키텍처

④ 훈련 과정은 한국수화언어 통역 모듈 훈련에 사용하지 않은 검증 데이터로 결과를 모니터링할 수 있도록 제작하였다. 한국수화언어 통역 모듈 Train batch 결과로 batch 반복 훈련 횟수별 훈련 데이터의 정확도와 검증 데이터의 정확도를 확인할 수 있고, <표 3-10>에 한국수화언어 통역 모듈 Train batch 결과의 예를 정리하였다.

<표 3-10> 한국수화언어 번역 모듈 Train batch 결과 예시

시간	한국수화언어 통역 Train batch 결과
2019-05-12 05:50:30,729	batch : 0000, loss : 2.30242, train_accuracy : 0.175781, test_accuracy : 0.17434
2019-05-12 05:52:18,343	batch : 0100, loss : 1.80412, train_accuracy : 0.560547, test_accuracy : 0.54605
2019-05-12 05:54:01,942	batch : 0200, loss : 1.12815, train_accuracy : 0.601562, test_accuracy : 0.62829
2019-05-12 05:55:41,685	batch : 0300, loss : 0.84690, train_accuracy : 0.802734, test_accuracy : 0.72697
2019-05-12 05:57:47,835	batch : 0400, loss : 0.46125, train_accuracy : 0.865234, test_accuracy : 0.81579
2019-05-12 05:59:41,907	batch : 0500, loss : 0.29823, train_accuracy : 0.878906, test_accuracy : 0.81250
...	
2019-05-12 07:35:46,770	batch : 5600, loss : 0.03785, train_accuracy : 0.998047, test_accuracy : 0.91447
2019-05-12 07:37:31,974	batch : 5700, loss : 0.03756, train_accuracy : 0.982422, test_accuracy : 0.90461
2019-05-12 07:39:23,308	batch : 5800, loss : 0.03762, train_accuracy : 0.978516, test_accuracy : 0.90789
2019-05-12 07:41:12,169	batch : 5900, loss : 0.09702, train_accuracy : 0.933594, test_accuracy : 0.86184
2019-05-12 07:43:12,686	batch : 6000, loss : 0.09274, train_accuracy : 0.986328, test_accuracy : 0.89803

⑤ 검증 모듈 제작 과정은 한국수화언어 통역 모듈을 로드하여 한국수화언어 특징 데이터 기반으로 사용 여부를 검증할 수 있는 모듈을 제작하였다. <표 3-11>은 한국수화언어 통역 모듈 검증 모듈로 한국수화언어 특징 JSON 데이터를 사용하여 한국수화언어 일상생활 대표 표제어로 예측한다.

<표 3-11> 한국수화언어 통역 모듈 검증 예시

입력 데이터(JSON)	[{'video_width':720, 'video_height':404, 'data': [{'KSL_id': 'KSL_10', 'prob': 45, 'left_x': 249, 'top_y': 244, 'width': 83, 'height': 90}]]]
특징맵 데이터	[0 0 0 ... 0 0 0]
특징맵 입력 Length	12,544개 (112×112)
통역 예측 모듈	prediction, prob = model.predict(x_input)
출력	prediction :: 6, prob :: 0.9969503283500671, vocab : 안녕



## (2) 한국수화언어 통역 서버 설계 및 구현

한국수화언어 통역 서버는 Flask[19]와 REST 기술[20]을 사용하여 제작되었다. Flask란 파이썬 웹 애플리케이션을 만드는 프레임워크로 매우 심플하고 가벼운 프레임워크이다. HTTP Method(POST, GET, PUT, DELETE)를 통해 해당 서버 접근이 가능하다. 제작 방법은 Flask를 사용하여 데이터 요청 기능을 제작하였고, 한국수화언어 특징 데이터를 JSON 포맷으로 한국수화언어 통역 모듈과 연동하여 한국수화언어 일상생활 대표 표제어로 예측하여 전송한다.

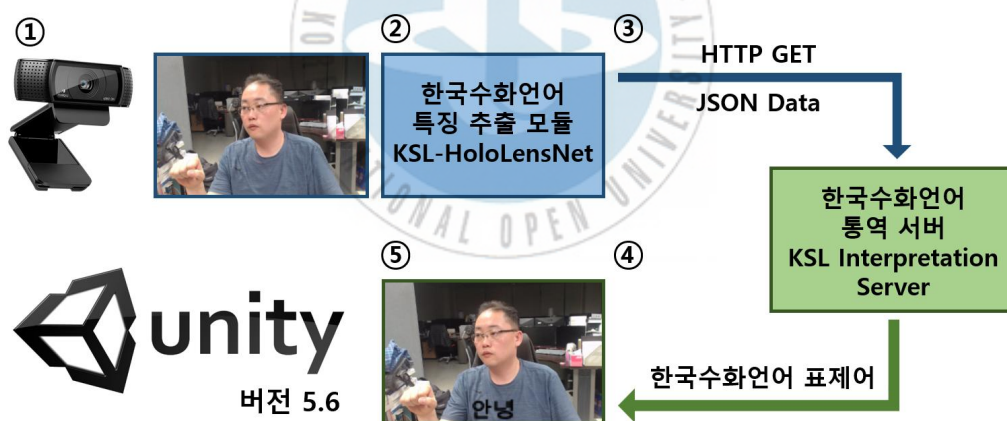
<표 3-12> 한국수화언어 통역 서버 적용 예시

한국수화언어 통역 서버	설명
* Serving Flask app "KSL_Interpretation_Server" (lazy loading)	한국수화언어 통역 서버 구동
* Environment: production	한국수화언어 통역 서버 환경
WARNING: Do not use the development server in a production environment.	경고 메시지
Use a production WSGI server instead.	WSGI(Web Server Gateway Interface) 사용 권유
* Debug mode: off	디버깅 모드
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)	한국수화언어 통역 서버 IP 및 포트
KSL_Interpretation_Server :: request :: [{ 'video_width': 720, 'video_height': 404, 'data':[{ 'KSL_id': 'KSL_10', 'prob': 45, 'left_x': 249, 'top_y': 244, 'width': 83, 'height': 90}, { 'KSL_id': 'KSL_45', 'prob': 24, 'left_x': 407, 'top_y': 234, 'width': 93, 'height': 117}]}]	한국수화언어 통역 요청
127.0.0.1 - - [14/May/2019 05:00:40] "prediction :: 6, prob :: 0.9969503283500671, vocab :: 안녕"	한국수화언어 통역 예측 결과

### 3.5. 한국수화언어 통역 App 설계 및 구현

#### (1) 한국수화언어 통역 App 설계

한국수화언어 통역 App은 한국수화언어 특징 데이터를 기반으로 한국수화언어 일상생활 대표 표제어로 통역하는 응용프로그램이다. 한국수화언어 통역 App은 유니티 5.6을 사용하였고, 증강/혼합현실용 HMD 웨어러블 디바이스에서 포팅이 가능하도록 설계되었다. <그림 3-10>과 같은 구동 프로세스를 갖는다. 한국수화언어 통역 프로세스는 ① 카메라의 영상 입력, ② 한국수화언어 특징 추출, ③ 한국수화언어 통역 서버에 HTTP GET 방식으로 한국수화언어 특징 JSON 데이터 전송, ④ 한국수화언어 표제어 예측, ⑤ 한국수화언어 일상생활 대표 표제어 출력의 프로세스 과정을 거친다.



<그림 3-10> 한국수화언어 통역 App 구동 프로세스 블록도



## (2) 한국수화언어 통역 App 구현

한국수화언어 통역 App 구현 절차는 <그림 3-11>과 같다. 한국수화언어 통역 App은 증강/혼합현실용 HMD 웨어러블 디바이스에서 구동할 수 있도록 Darknet[22]을 DLL로 포팅하였다. 증강/혼합현실용 HMD인 홀로렌즈의 경우 윈도우10 32bit OS 환경이기 때문에 Darknet을 x86 설정으로 yolo\_cpp\_dll의 이름으로 DLL 컴파일하여 포팅하였다.

한국수화언어 통역 App 세부 구동 과정은 다음과 같다.

- ① 유니티 엔진과 Darknet 연동 과정은 마샬링 기법을 사용하였다.
- ② 한국수화언어 특징 추출 모듈 가중치 파일과 환경설정 파일 로딩 과정은 KSL-HoloLensNet 가중치 파일(weights)과 환경설정 파일(cfg)을 로딩하였다.
- ③ 한국수화언어 특징 추출 모듈은 OpencvForUnity를 사용하여 영상 프레임별로 Darknet KSL-HoloLensNet 모듈에 이미지 데이터를 byte[] 포맷으로 전달하여 한국수화언어 특징을 검출한다.
- ④ 한국수화언어 통역 요청 과정으로 한국수화언어 통역 서버로 한국수화언어 특징 JSON 데이터를 전달하여 한국수화언어 일상생활 대표 표제어를 요청한다.
- ⑤ 한국수화언어 일상생활 대표 표제어 출력 과정으로 한국수화언어 통역 서버에서 한국수화언어 일상생활 대표 표제어를 정보를 받아 출력한다.

홀로렌즈와 같은 증강/혼합현실용 HMD 웨어러블 디바이스 빌드는 유니티 Windows Store 빌드와 홀로렌즈 앱 빌드를 추가로 진행하여 빌드할 수 있다.



<그림 3-11> 한국수화언어 통역 App 구현



## 제4장 실험

### 4.1. 한국수화언어 특징 추출 모듈 실험 및 평가

한국수화언어 특징 추출 모듈 평가는 실시간 객체 검출 모듈인 yolov3-tiny 모듈에 한국수화언어 특징 추출 적용한 KSL-yolov3-tiny 모듈과 증강/혼합현실용 웨어러블 디바이스 구동 목적으로 제안한 KSL-HoloLensNet 모듈을 비교 평가하였다. 평가 대상은 KSL-yolov3-tiny 모듈과 제안한 KSL-HoloLensNet 모듈이다. 한국수화언어 특징 추출 모듈 평가 항목은 검출 속도 측면과 정확도 측면으로 나누어 평가한다. 한국수화언어 특징 추출 모듈 검출 속도 측면 평가 항목은 연산량 BFLOPS와 PC 환경에서 GPU를 사용한 경우와 CPU를 사용한 경우로 나누어 측정한 한국수화언어 검출 속도이다. 한국수화언어 특징 추출 모듈 정확도 측면 평가는 훈련에 사용되지 않은 검증 데이터셋 768장을 사용하여 평균정밀도 mAP, 조화평균값 F1-score, 정밀도 Precision, 재현율 Recall, 평균 IoU를 평가한다. 마지막으로, 모듈 크기를 평가한다.

한국수화언어 특징 추출 모듈 평가 항목 및 실험 결과를 <표 4-1>에 정리하였다. 한국수화언어 특징 추출 모듈 평가 항목 중 검출 속도 측면에서 KSL-yolov3-tiny 모듈과 KSL-HoloLensNet 모듈의 BFLOPS를 비교하면 5.557와 0.615으로 제안 모델이 약 1/9의 연산만으로 특징 추출이 가능하다. GPU를 사용한 검출 시간은 5ms와 1ms로 제안 모델이 빠른 검출 성능을 갖는다. 증강/혼합현실용 HMD 웨어러블 디바이스에서는 GPU를 사용할 수 없는 상황이 많으며, 이때 CPU 연산으로 처리해야 한다. CPU를 사용한 검출 속도는 각각 255ms와 75ms로 제안 모델이 3.4배 빠른 성능을 갖는다. 멀티스레드 기법을 적용하면 증강/혼합현실용 HMD 웨어러블 디바이스에서도 빠른 성능을 기대할 수 있다. 한국수화언어 특징 추출 모듈 평가 항목 중 정확도 측면인 평균정밀도 mAP은 95.84%와 94.02%로 제안 모델에서 1.82% 감소하였고, 조화평균값 F1-score는 91%와 87%로 제안 모델에서 4% 감소하였다. 마지막으로 모듈 크기는 34.52mb와 6.92mb로 제안 모델이 27.6mb 작게 경량화 되었다. 따라서, 제안한

KSL-HoloLensNet 모듈은 KSL-yolov3-tiny 모듈과 비교하여 연산량은 적고, 정확도는 유지할 수 있는 성능의 경량화된 모듈이라 할 수 있다. 이러한 모듈 성능은 증강/혼합현실용 HMD와 같은 웨어러블 플랫폼에서 적용 가능 수준이라 평가된다.

<표 4-1> 한국수화언어 특징 추출 모듈 평가

평가 항목	KSL-yolov3-tiny	KSL-HoloLensNet
Test Data	768개	768개
Total BFLOPS	5.557	0.615
Input Image Size	416	352
Threshold	0.25	0.25
Total Detection Time (768개)	17초	11초
GPU Detection Time (1개)	5ms	1ms
CPU Detection Time (1개)	255ms	75ms
mean AveragePrecision(mAP)	95.84%	94.02%
F1-score	91%	87%
Precision	90%	81%
Recall	93%	93%
average IoU	75.97%	61.24%
Module Size	34.52MB	6.92MB

## 4.2. 한국수화언어 통역 모듈 실험 및 평가

한국수화언어 통역 모듈은 KSL\_NLP\_dataset\_yolov3 데이터셋으로 훈련한 KSL\_NLP\_dataset\_yolov3 통역 모듈과 KSL\_NLP\_dataset 데이터셋으로 훈련한 KSL\_NLP\_dataset 통역 모듈을 비교하여 평가하였다. 한국수화언어 통역 모듈의 정확도는 한국수화언어 일상생활 대표 표제어를 기반으로 훈련에 사용되지 않은 평가 데이터로 평가하였다. 정확도는 한국수화언어 대표 표제어 예측치 오차를 평균치로 표시하였다.

한국수화언어 통역 모듈의 평가 항목과 결과는 <표 4-2>에 정리하였다.

KSL\_NLP\_dataset 통역 모듈은 훈련 데이터 15,842개, 평가 데이터 1,760개로 한국수화언어 일상대화 대표 표제어 10개 극성을 가지며, 정확도는 95.85%로 나타났다. KSL\_NLP\_dataset\_yolov3 통역 모듈은 훈련 데이터 2,743개, 평가 데이터 304개로 한국수화언어 일상대화 대표 표제어 10개 극성을 가지며, 정확도는 91.44%로 나타났다.

한국수화언어 통역 모듈은 일상생활 대표 표제어 10개 기준으로 제작되어 추후 표제어 확장 연구가 필요하다.

<표 4-2> 한국수화언어 통역 모듈 평가

평가 항목	KSL_NLP_dataset_yolov3 통역 모듈	KSL_NLP_dataset 통역 모듈
데이터셋 아이템 수	112개	464개
전처리 Input Size (한국수화언어 특징맵)	112×112	112×112
Train Data	2,743개	15,842개
Test Data	304개	1,760개
표제어 수	10개	10개
정확도	91.44%	95.85%

### 4.3. 한국수화언어 통역 App 실험 및 평가

한국수화언어 통역 App 실험 및 평가는 윈도우 10 OS, AMD Ryzen 7 2700 Eight-Core Processor 3.20 GHz, 16GB RAM, Nvidia RTX-2080TI 환경에서 테스트 되었다. 증강/혼합현실용 HMD 웨어러블 디바이스에서는 GPU를 사용할 수 없는 환경이 많아 CPU 연산을 사용하여 실험을 진행하였다.

한국수화언어 통역 App은 한국수화언어 특징 추출 KSL-HoloLensNet 모듈과 한국수화언어 통역 KSL\_NLP\_dataset 모듈을 사용하여 <그림 4-1>과 같이 구동하였다.

한국수화언어 통역 App의 평가 항목은 한 프레임의 한국수화언어 통역 소요 시간과 FPS를 비교하고, 싱글스레드 기법과 멀티스레드 기법 적용에 따라 평가하였다. 평가 항목 및 평가 결과는 <표 4-3>에 정리하였다. 한국수화언어 통역 App에서 멀티

스레드 기법을 적용하면 CPU 연산을 사용하여 한국수화언어 통역에서 출력까지 25ms 소요되며, 28 FPS의 성능을 나타내었다. 이는 실시간으로 한국수화언어 통역이 가능한 성능이라 평가된다. 향후 연구를 통해 증강/혼합현실용 웨어러블 디바이스에 서도 실시간 한국수화언어 통역 시스템 적용이 가능할 것으로 판단된다.



<그림 4-1> 한국수화언어 통역 App 구동 화면

<표 4-2> 한국수화언어 통역 App 평가

평가 항목	싱글스레드 기법	멀티스레드 기법
한국수화언어 통역 소요 시간	232 ms	25 ms
FPS(Frames Per Second)	4 FPS	28 FPS



#### 4.4. 한국수화언어 통역 시스템 고찰

한국수화언어 통역 시스템은 실시간으로 한국수화언어 통역이 가능한 시스템이다. 한국수화언어 통역 App은 CPU 환경에서 멀티스레드 기법을 적용하면 28 FPS의 성능을 갖는다. 정확도 측면에서는 한국수화언어 특징 추출 모듈 KSL-HoloLensNet은 한국수화언어 10개 표제어의 평균정밀도는 mean AveragePrecision(mAP) 94.02%이고, 조화평균값 F1-score는 87%의 성능이다. 한국수화언어 통역 모듈인 KSL\_NLP\_dataset 모듈은 95.85% 정확도의 성능이다.

한국수화언어 통역 시스템은 한국수화언어 일상생활 대표 표제어 10개라는 동일한 조건의 기존 연구[9]의 정확도 84%보다 높은 성능을 보였다. 특히 본 논문에서 제한한 한국수화언어 통역 시스템은 실시간으로 한국수화언어 통역이 가능하다는 점에서 기존 연구와 차이가 있다.

한국수화언어 통역 시스템은 기존 연구에서 사용된 데이터 클러브[13], 키넥트[14], 립모션[15] 등의 특수장비에 종속되지 않는 특징을 갖는다. 립모션과 같은 수지 요소만을 검출하여 통역하는 시스템과 달리 비수지 요소를 포함할 수 있다는 장점이 있다. 또한 한국수화언어 통역 시스템은 저비용의 카메라로 촬영된 영상으로 데이터를 수집이 가능하다.

한국수화언어 통역 시스템 연구의 제한점은 다음과 같다. 첫째, 한국수화언어 통역 시스템은 한 화자의 수화언어만 통역이 가능하다. 이는 딥러닝을 이용한 모듈 훈련을 한 화자의 수화언어를 기반으로 제작되었기 때문이다. 둘째, 한국수화언어 통역 시스템은 손 모양의 수형을 65가지로 코드화하여 사용한다. 이는 손 모양의 수형에 따라 수형코드로 검출할 수 있으나 동일한 수형에 있어 회전과 같은 시퀀스 정보 포함할 수 없는 제한점이 있다.

한국수화언어 통역 시스템의 향후 연구 과제는 다음과 같다. 첫째, 한국수화언어 통역 시스템은 한국수화언어 일상생활 대표 표제어 10개로 제작되어 표제어 개수가 작아 표제어 확장 연구가 필요하다. 둘째, 한국수화언어 통역 시스템의 한국수화언어 특징 추출 모듈의 비수지 요소인 얼굴 요소의 확장이 필요하다. 한국수화언어 특징 추출 모듈의 비수지 요소는 기본 웃는 얼굴과 ‘아’, ‘에’, ‘이’, ‘오’, ‘우’를 코드화하



여 사용하였다. 비수지 요소인 얼굴의 감정 표현에 따라 한국수화언어 의미가 달라질 수 있기 때문에 연구가 필요하다. 셋째, 한국수화언어 통역 시스템의 지문자 통역 확장 연구가 필요하다. 한국수화언어 통역 시스템은 자연수화언어 동작(단어)에 대하여 연구되어 지문자 통역에 관한 연구가 필요하다.

한국수화언어 통역 시스템은 제한점과 향후 연구 과제가 있지만, 다음과 같은 연구 가치를 갖는다. 첫째, 한국수화언어 통역 시스템은 실시간 통역 성능을 가진다. 둘째, 특수장비가 필요 없고 GPU와 같은 고성능 하드웨어를 필요로하지 않는다. 셋째, 한국수화언어 훈련 데이터 수집이 용이하다. 마지막으로 PC 및 증강/혼합현실용 HMD와 같은 웨어러블 디바이스와 모바일 디바이스에 포팅이 가능하다.



## 제5장 결론

본 논문에서는 한국수화언어 통역 시스템을 제안하였다. 한국수화언어 통역 시스템은 한국수화언어 특징 추출 모듈과 한국수화언어 통역 모델로 구성되어 있다. 한국수화언어 통역 모듈은 한국수화언어 일상생활 대표 표제어 10개를 중심으로 개발되어 실제 사용에는 제한이 따른다. 향후 한국수화언어 표제어 확장 연구를 통해 한국수화언어의 다양한 자연어처리가 가능할 것으로 생각된다.

본 논문을 통해 개발된 한국수화언어 통역 시스템은 한국수어 사전 통역 서비스로 사용할 수 있다. 또한, 공공부문 활용에 응용될 수 있다. 한국어를 사용하는 일반이 한국수화언어 사용자와 소통을 위해서는 한국수화언어를 학습해야 하는 어려움이 있어, 한국수화언어를 사용하는 농인 및 청인은 공공기관에서 한국수화언어로 서비스를 받기 어려운 것이 현실이다. 한국수화언어 통역 시스템은 주민 센터, 법원 등 공공기관에서 한국수화언어 통역 시스템을 활용하여 서비스를 받을 수 있을 것으로 기대된다.

본 논문에서 제안한 한국수화언어 통역 시스템은 데이터 글로브나 키넥트, 립모션과 같은 특수장비를 사용하지 않고, 카메라를 통해 입력된 영상을 사용한다. 따라서 PC 환경 및 증강/혼합현실 HMD와 같은 웨어러블 디바이스뿐만 아니라 휴대폰과 같은 범용적인 모바일 디바이스 환경에서 활용할 수 있다.

마지막으로 한국수화언어 대화 가이드 시스템으로 확장할 수 있다. 증강/혼합현실용 HMD 디바이스에서 연속된 한국수화언어 자연어처리 기술을 사용하여, 시스템 사용자 음성 인식 후 표현하고자 하는 한국수화언어를 안내를 받을 수 있는 한국수화언어 대화 가이드 시스템으로 확장할 수 있다.

## 참고문헌

- [1] 국가법령정보센터, (2016년 8월 4일), 한국수화언어법, 주소 : <http://www.law.go.kr/%EB%B2%95%EB%A0%B9/%ED%95%9C%EA%B5%AD%EC%88%98%ED%99%94%EC%96%B8%EC%96%B4%EB%B2%95> (2019년 4월 10일 다운로드)
- [2] 이슬기, (2013년 6월 13일), “‘수화’ VS ‘수어’ 끊이지 않는 논쟁”, 주소 : <http://www.ablenews.co.kr/News/NewsContent.aspx?CategoryCode=0014&NewsCode=001420130603175808498076> (2019년 5월 13일 다운로드)
- [3] 강혜민, (2013년 10월 22일), “‘언어로서의 수화’인 ‘한국수어’ 법안 발의돼”, 주소 : <http://beminor.com/detail.php?number=6021> (2019년 5월 13일 다운로드)
- [4] 구글, (2014년 7월 3일), 수화 자동 인식 스마트 글라스 장치, 주소 : <https://patents.google.com/patent/KR20160004561A/ko> (2019년 4월 10일 다운로드)
- [5] 구글, (2017년 8월 3일), 지화 인식 시스템 및 방법, 주소 : <https://patents.google.com/patent/WO2017131318A1/ko> (2019년 4월 10일 다운로드)
- [6] 고광은, 심귀보 “딥러닝을 이용한 객체 인식 및 검출 기술 동향”, 제어로봇시스템학회지 23(3), pp. 17-24, 2017년..
- [7] 이준우, 남기현, “한국 수어학 개론”, 나남, 2014년
- [8] 임재철, 김광현, 반영환, “스마트 장갑을 이용한 수화 인식 인터랙션 시스템 연구”, 한국디자인학회 학술발표대회 논문집, 98 - 99, 2015.10
- [9] 신효주, “CNN 기반 노이즈가 포함된 한국 수화 동작 인식에 관한 연구”, 서울과학기술대학교, 2018년
- [10] 이준우, “수형 기반 한국수어사전 구축”, 국립국어원, 2016년
- [11] 이병환, 이기성 “PC 카메라에서 추출한 이미지를 이용한 수화인식” 정보 및 제어 논문집, 102-104, 2004년
- [12] 양희덕, “Conditional Random Field와 BoostMap 임베딩 기반 수지 및 비수지 신호를 이용한 수화 인식” 정보과학회논문지 : 소프트웨어 및 응용, 41(1), 96-101, 2014년

- [13] David J. Hill, (2012, Sep 16), Smart Gloves Turn Sign Language Gestures Into Vocalized Speech, Available : <http://singularityhub.com/2012/09/16/smart-gloves-turn-sign-language-gestures-into-vocalized-speech/> (downloaded 2019, May. 10)
- [14] Michael Kan, (2013년 11월 1일), 수화 번역기로 변신한 MS 키넥트 ... 청각 장애인 위한 ‘키넥트 번역기’ 나온다, 주소 : <http://www.itworld.co.kr/news/84445> (2019년 5월 13일 다운로드)
- [15] 전빈, 권재영, 박준규, 이승훈, 이희연, 장현국 “립모션을 활용한 웨어러블 수화 통역기” 한국정보과학회 학술발표논문집, 1461-1463, 2018년
- [16] Garcia, Brandon, and Sigberto Alarcon Viesca. “Real-time american sign language recognition with convolutional neural networks.” Convolutional Neural Networks for Visual Recognition 2, 2016년
- [17] wikipedia, (2019년 4월 4일), 딥 러닝, 주소 : [https://ko.wikipedia.org/wiki/%EB%94%A5\\_%EB%9F%AC%EB%8B%9D](https://ko.wikipedia.org/wiki/%EB%94%A5_%EB%9F%AC%EB%8B%9D), (2019년 4월 10일 다운로드)
- [18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587, 2014
- [19] R. Girshick, “Fast r-cnn,” Proceedings of the IEEE International Conference on Computer Vision, pp. 1440-1448, 2015
- [20] Redmon, Joseph; Divvala, Santosh; Girshick, Ross; Farhadi, Ali, “You Only Look Once: Unified, Real-Time Object Detection”, eprint arXiv:1506.02640, 06. 2015
- [21] W. Liu et al., “Ssd: Single shot multibox detector”, European conference on computer vision, pp. 21-37, 2016
- [22] AlexeyAB, (2017, Nov. 28), Yolo-v3 and Yolo-v2 for Windows and Linux, Available : <https://github.com/AlexeyAB/darknet> (downloaded 2019, April. 10)
- [23] Iandola, Forrest N., et al. “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size.” arXiv preprint arXiv:1602.07360 (2016).
- [24] Huang, Gao, et al. “Densely connected convolutional networks.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

- [25] Microsoft, (2018, Mar. 21), HoloLens hardware details, Available : <https://docs.microsoft.com/en-us/windows/mixed-reality/hololens-hardware-details> (downloaded 2019, April. 10)
- [26] MindUOwn, (2016, Mar. 5), HoloLens Hardware Specifications, Available : <http://www.vrtalk.com/forum/showthread.php?2047-HoloLens-Specifications-Released> (downloaded 2019, April. 10)
- [27] 구민재, 이병래, 조동빈, “딥러닝을 이용한 실시간 손 검출기: 증강현실 HMD(Head Mounted Display) 홀로렌즈 적용을 중심으로”, JCCI 2019
- [28] Shi Qiu, (2017, Sep. 12), A simple tool for labeling object bounding boxes in images, Available : <https://github.com/puzzledqs/BBox-Label-Tool> (downloaded 2019, April. 10)
- [29] 국립국어원, (2019년), 한국수어 사전, 주소 : <http://sldict.korean.go.kr/front/main/main.do> (2019년 3월 7일 다운로드)
- [30] taki0112, (2019, Mar. 4), Densenet-Tensorflow, Available : <https://github.com/taki0112/Densenet-Tensorflow> (downloaded 2019, May. 13)
- [31] Armin Ronacher, (2019), Flask, Available : <http://flask.pocoo.org/> (downloaded 2019, May. 13)
- [32] wikipedia, (2018, Sep. 15), REST, Available : <https://ko.wikipedia.org/wiki/REST> (downloaded 2019, May. 13)

# A B S T R A C T

## Design and Implementation of Korean Sign Language Interpretation System Using Deep Learning

by

***Koo, Min Jae***

Department of Public Administration  
Graduate School  
Korea National Open University

Supervised by Professor : Lee, Byeong Rae

The Korean Sign Language is defined as the language of the Korean Language equivalent to that of the Korean Language, and aims to enhance the language of the deaf and improve the quality of life. There is no difficulty in communicating using the Korean Sign Language. However, it is difficult for the general person to learn Korean Sign Language to communicate with the deaf and the hearer who use Korean Sign Language and the public using Korean Language. In order to improve the communication problem, we would like to propose a Korean Sign Language Interpretation System Using Deep Learning to understand the meaning of Korean Sign Language.

The Korean Sign Language Interpretation System consists of the Korean Sign Language Feature Extraction Module and the Korean Sign Language Interpretation Module. The Korean Sign Language Feature Extraction Module is an object detection module that extracts the type and face type probability and position information of hand and finger



shape. The Korean Sign Language Feature Extraction Module can detect and classify Korean Sign Language features into 71 classes. The calculation amount is 0.615 BFLOPS and the object detection time using CPU operation is 75ms. The accuracy of the Korean Sign Language Feature Extraction Module is 94.02% for mean average precision (mAP), 87% for F1-score, and 61.24% for average IOU. Finally, the Korean Sign Language Feature Extraction module size is 6.92MB, which is a lightweight module.

The Korean Sign Language Interpretation Module is a module that interprets the natural sign language behavior (words) among the Korean Sign Language based on the information extracted using the Korean Sign Language Feature Extraction Module. The Korean Sign Language Interpretation Module was produced for 10 representative headlines of Korean Sign Language, and the accuracy of the Korean Sign Language Interpreter Module was 95.85% when 1,760 evaluation data were not used for training. The Korean Sign Language Interpreter App has a performance of 28 FPS, which takes 25ms from the Korean Sign Language Interpretation to the output when the CPU operation is applied using the multi-thread technique. This is evaluated as the ability to interpret Korean Sign Language in realtime.

The study value of the Korean Sign Language Interpretation System Using Deep Learning is as follows. First, the Korean Sign Language Interpretation System has realtime interpretation performance. Second, it does not require special equipment and does not require high-performance hardware such as a GPU. Third, it is easy to collect Korean Sign Language training data. Finally, it is possible to port to wearable devices and mobile devices such as PC and HMD for augmented/mixed reality.

## 감사의 글

2017년 한국방송통신대학교 정보과학에 입학하고 5학기가 지났습니다. 디자인학과 출신, 게임 개발자, 남들하고 조금은 다른 이력으로 학업을 시작하였고 인공지능이란 재미난 학문을 접하게 되었습니다. 저의 연구가 조금은 사회에 유익이 되는 곳에 쓰일 수 있도록 딥러닝을 이용한 한국수화언어 통역 시스템 설계 및 구현을 작성하게 되었습니다.

본 논문과 관련하여 1년이 넘는 기간 동안 지도해주신 이병래 교수님께 감사를 드립니다. 정기적으로 연구 결과에 대한 지도를 해주셨고, 특히 2019년 2월에 JCCI 2019 논문을 제출하는 과정에서도 논문에 대한 많은 조언을 해주셔서 감사드립니다.

본 논문을 심사해주신 손진곤, 이관용 교수님께도 감사를 드립니다. 제 논문의 부족한 부분들을 지적해주시고 보완 방향에 대한 다양한 의견을 주셔서, 더 완성도 높은 논문을 작성할 수 있었습니다. 튜터와 조교님들에게 감사를 드립니다. 수업 과정에서 여러 질의와 요청을 드리고, 학사 행정과 관련된 많은 질의에도 친절하고 성실하게 답변해주셔서, 원활한 대학원 생활을 마칠 수 있었습니다.

동기분들과 여러 선배님에게도 감사를 드립니다. 2년 넘는 과정 동안 제가 지치고 힘들어할 때 많은 도움과 힘을 주셔서 대학원 과정을 잘 마칠 수 있었습니다. 한양스터디를 이끌어 가시는 원우님들의 노고에 감사를 드립니다. 꾸준히 이어온 스터디 덕분에 대학원 과정의 어려움을 슬기롭게 극복할 수 있었습니다.

마지막으로 우리 가족들에게 감사를 표합니다. 아들 시원 그리고 아내 이주영에게 진심으로 감사를 표합니다.

# 학위논문 공개 동의서

한국방송통신대학교 대학원은 귀하의 학위논문을 데이터베이스화하여 내용의 일부 또는 전부를 어떤 형태로든 국내·외 이용자에게 공개(열람 또는 배포)하는 것에 대하여 동의를 얻고자 하오니 동의여부를 기재하여 주시기 바랍니다.

학 위	이학석사		졸업년도	2019
전 공	정보과학과			
학위논문 제 목	한글	딥러닝을 이용한 한국수화언어 통역 시스템 설 계 및 구현		
	영문	Design and Implementation of Korean Sign Language Interpretation System Using Deep Learning		
동 의 여 부	1. 찬성( ○ )    2. 조건부 찬성(   )    3. 반대(   )			
조건 또는 사유	※ “조건부 찬성” 또는 “반대”하는 경우 그 사유 기재			

2019 년 07 월 23 일

논문저자명 : 구 민 재 (서명 또는 날인)

**한국방송통신대학교총장 귀하**