

## Development of A Uniform And Casual Clothing Recognition System For Patient Care In Nursing Hospitals

Ye-Chan Yun\*, Young-Tae Kwak\*

\*Student, Dept. of Information Technology & Engineering, Jeonbuk National University, Jeonju, Korea

\*Professor, Dept. of Information Technology & Engineering, Jeonbuk National University, Jeonju, Korea

### [Abstract]

The purpose of this paper is to reduce the ratio of the patient accidents that may occur in nursing hospitals. In other words, it determines whether the person approaching the dangerous area is a elderly (patient uniform) group or a practitioner(Casual Clothing) group, based on the clothing displayed by CCTV. We collected the basic learning data from web crawling techniques and nursing hospitals. Then model training data was created with Image Generator and Labeling program. Due to the limited performance of CCTV, it is difficult to create a good model with both high accuracy and speed. Therefore, we implemented the ResNet model with relatively excellent accuracy and the YOLO3 model with relatively excellent speed. Then we wanted to allow nursing hospitals to choose a model that they wanted. As a result of the study, we implemented a model that can distinguish patient and casual clothes with appropriate accuracy. Therefore, it is believed that it will contribute to the reduction of safety accidents in nursing hospitals by preventing the elderly from accessing the danger zone.

▶ **Key words:** Nursing Hospital, Patient Uniform, Casual Clothing, Clothing Recognition

### [요 약]

본 연구의 목적은 요양병원에서 발생할 수 있는 노인안전사고 발생률을 감소시키는 것이다. 즉, 위험지역으로 접근하는 인물이 노인(환자복) 그룹인지 실무자(평상복) 그룹인지를 CCTV에 나타나는 의복을 기준으로 구별하는 것이다. Web Crawling 기법과 요양병원으로부터 지원을 받아 기초 데이터를 수집하였다. 이후 Image Generator와 Labeling으로 모델 학습 데이터를 만들었다. CCTV의 제한된 성능 때문에 높은 정확도와 속도를 모두 갖춘 모델을 만드는 것은 어려웠다. 그러므로 정확성이 상대적으로 우수한 ResNet 모델, 속도에서 상대적으로 우수한 YOLO3 모델을 각각 구현했다. 그리고 요양병원이 자신의 실정에 맞는 모델을 고를 수 있게 하고자 했다. 연구 결과 환자복과 평상복을 적절한 정확도로 구별할 수 있는 모델을 구현하였다. 따라서 실제 사용처에서 노인들이 위험구역에 접근하지 못하도록 하여 요양병원 안전사고 감소에 이바지 할 것으로 평가된다.

▶ **주제어:** 요양병원, 환자복, 평상복, 의복 인식

- 
- First Author: Ye-Chan Yun, Corresponding Author: Young-Tae Kwak
  - \*Ye-Chan Yun (yunyc1020@naver.com), Dept. of Information Technology & Engineering, Jeonbuk National University
  - \*Young-Tae Kwak (ytkwak@jbnu.ac.kr), Dept. of Information Technology & Engineering, Jeonbuk National University
  - Received: 2020. 10. 12, Revised: 2020. 11. 05, Accepted: 2020. 12. 09.

## I. Introduction

대한민국 사회에서 노인인구 증가는 세계에서 가장 빠르게 진행되고 있다. 2000년도에 65세 이상 인구가 전체 인구의 7%를 돌파하게 되며 고령화 사회가 되었고, 2017년에 14%를 돌파하며 대한민국은 고령사회로 진입하게 되었다[1]. 그리고 현재 2020년 기준, 대한민국 65세 이상 고령 인구 비율은 16%이다[2].

이러한 노인인구 증가속도와 더불어 경제수준 향상 등으로 대한민국 노인들의 생활수준이 전반적으로 높아지면서, 의식주 문제를 넘어선 다양한 욕구해결을 위한 노인 관련 서비스에 대한 수요가 증가하게 되었다[3]. 특히 요양병원과 같은 건강관련 시설의 확대를 요구하는 목소리는 시간이 지날수록 확대되는 모습을 보여주었고, 실제로 이 목소리를 반영하듯 요양병원과 요양원은 2017년 기준, 작년 대비 15% 증가하며 총 6883개소로 나타났다[4][5].

노인인구 증가와 요양병원시설 증가 및 확대로 노인보호 관련 문제와 안전사고가 증가하게 되면서 이와 관련된 사안들이 우리사회에서 해결해야 할 중요한 안건으로 부상하게 되었다. 이런 안전사고의 대표적인 원인으로 치매노인의 증가와 노령화로 인한 노인의 인지능력저하로 인해 노인과 간호사간의 의사소통에 장애가 발생하면서 관계자 외 출입금지 지역과 위험지역으로의 노인들의 접근이 증가하게 되었으며 이러한 접근을 관계자 및 간호사와 같은 전문 인력들이 모두 막을 수 없음을 꼽을 수 있다[6][7].

이에 본 연구는 전문 인력들이 노인들께서 위험지역에 접근하지 못하도록 24시간 감시할 수 없다는 현실적인 실정과 함께 또 이에 준하는 노력을 함에도 불구하고 안전사고가 발생할 수 밖에 없는 현 상황을 고려하여 노인 안전사고 발생률을 최소화하기 위한 기술을 제안하고자 한다.

본 연구의 핵심은 인공지능과 컴퓨터비전 기술을 결합하여 관계자 외 출입금지 지역, 위험지역으로 접근하는 인물을 CCTV로 확인하고, 그 인물이 입고 있는 옷을 스캔한 뒤, 미리 학습한 신경망을 바탕으로 관계자(평상복)그룹인지 노인(환자복)그룹인지를 구별하는 것이다.

두 그룹을 구별하기 위하여 사용한 기술로는 CNN구조에서 정확성을 더 향상시킨 ResNet-50모델과 Real-Time Object Detection을 지원하는 YOLO3모델을 사용하였다. 먼저 이미지를 바탕으로 지도학습을 수행하여 두 그룹을 구별할 수 있는 신경망을 구축했고, 이후 실제 CCTV에 본 연구에서 개발된 신경망을 적용하여 실시간으로 영상에서 인물이라는 객체를 탐지, 그룹 구별을 수행할 수 있도록 하였다.

본 연구를 수행하게 된 배경은 노인안전사고 감소라는 큰 목표 외에도 두 가지가 더 있다. 첫째, 요양병원 입장에서 볼 때 소비자들에게 더 나은 의료서비스를 제공시켜 드릴 수 있으므로 소비자의 서비스 재이용 의도를 높일 수 있을 것이다[8]. 둘째, 물리보안시장이 활성화되고, AI 시장이 발전되고 있는 현 시국에 맞는 기술개발은 요양병원 외에도 더욱 많은 시설에 도움이 될 것이라는 점이다[9].

지금부터 이러한 연구방법과 목적, 그리고 배경을 가지고 본 연구가 목표로 했던 요양병원에서 환자들이 허가 되지 않는 공간으로 출입하는 문제를 어떻게 해결하고자 하였는지 상세하게 설명하고, 이후 본 연구 결과가 지닌 의의와 한계점을 제시하도록 하겠다.

## II. Preliminaries

### 1. Tracking a Specific Person Using Clothes Color

의상색상을 이용하여 특정인을 추적하는 기술은 인간처럼 행동하는 로봇기술과 CCTV 응용기술에서 많이 등장하였고, 현재도 널리 통용되고 있는 기술이다[10]. 그러나 로봇과 CCTV는 특정인 추적 기술의 적용 방법이 다르다. 로봇은 센서를 통해 실시간으로 색상을 감지하고 스스로 움직여 대상추적을 위한 데이터를 수집할 수 있지만, CCTV는 획득한 영상만을 가지고 특정인을 추적하여야 한다. CCTV의 가장 큰 문제점은 시스템 속도인데 이 문제를 해결하기 위해서는 입력된 영상을 계산에 사용하기 전에 불필요한 데이터는 생략하고 정제하는 과정이 필요하며, 계산 수행 중에는 연산 수를 최소화하며 계산을 수행해주어야 한다. 제한된 시스템 속도에서 높은 정확도와 인식률을 달성하고자 많은 연구들이 선행되어 왔는데 그 중 대표적인 몇 가지의 예를 들자면, 기존의 방법이라 할지라도 속도가 상대적으로 빠른 라이브러리나 API를 사용하는 방법과 기존의 계산식을 연구나 개발이 적용될 환경과 상황, 그리고 조건을 고려하여 새로운 계산식으로 변경해주는 방법이 있다.

### 2. Fashion Image Searching Website based on Image Classification

의상인식은 단순히 의상구별과 객체 인식에만 사용되는 것이 아닌 획득한 의상 이미지를 분류하고 답러닝으로 학습시키는 과정을 거쳐 다양한 응용 분야에 접목시킬 수 있는 연구 분야이다. 한 가지 예시로 기존에 존재하는 패션

웹 사이트에서는 상의, 하의 등의 품목에서 한 가지 의류만의 정보를 제공하기 때문에 사용자가 원하는 옷에 대한 조합을 찾을 수 없다. 이러한 문제를 해결하기 위해 딥러닝(주로 CNN 모델을 사용)을 통한 이미지 분류를 웹 사이트와 연동하는 아이디어를 고려해볼 수 있다. 웹 사이트에 사용자가 본인의 이미지를 업로드 하면 딥러닝 서버를 통해서 이미지의 특징들을 추출하고 이 특징을 분류하여 저장한다. 그럼 사용자는 저장된 데이터를 가지고 여러 조합을 통해 원하는 이미지를 검색 할 수 있다[11].

### 3. Virtual-Try-On

가상착용기술(Virtual-Try-On: VTON)이란 사용자가 직접 의류나 장신구를 착용하지 않고도 인물의 이미지를 입력으로 받아 인물의 자세 및 체형, 의상의 가려짐 정도, 의상의 특성 등을 고려하여 인물의 이미지에 의류나 장신구를 합성시켜주는 기술을 말한다. 최근 활용도가 매우 높아 IT분야에서 핵심기술로 부상하고 있는 AR과 VR시장의 성장세에 발맞추며 같이 발전되어 가고 있는 연구 분야 중 하나이다. 현재는 이 분야에서 가장 좋은 성능을 보이는 CP-VTON 딥러닝 모델에서 몇 가지 문제점이 검출되었다. 심각한 문제점의 하나로 인물의 Neck/Check 부분이 백그라운드로 구분됨으로써, 의상의 목 부분이 인물의 현재 의상에 영향을 받게 되는 점을 들 수 있다. 이 문제를 해결하는 방법으로 목 부분의 Label인 Skin을 추가하고, 또한 머리카락 등 불필요 데이터를 제외하여 사용자 실루엣을 구성하는 등의 방법이 제시되었다. 이 외에도 여러 문제점을 내포하고 있으며 현재도 추가적인 연구를 통해 문제점들이 개선되어가고 있다[12].

### 4. Visual Fashion Analysis

시각적 의류 분석 기술은 대량의 의류 이미지의 분석과 관리에 대한 필요성이 산업의 발달에 따라 커지면서 각광 받게 된 기술이다. 이 기술에는 의류의 종류를 구별하는 의류 인식과 비슷하거나 같은 의류를 찾는 의류 검색 등의 기술이 포함된다[13]. 의류 인식이란 영상 속에 있는 의류를 추출하고, 추출한 의류에서 특성정보(plaid, flare, button)를 바탕으로 의류의 분류를 결정하는 것을 말한다. 의류 인식에는 FashionNet, Attentive Fashion Grammar Network 등의 방법이 있다. 의류 검색이란 입력된 이미지를 표준 이미지들과 비교하여 매칭시키는 작업이다. 의류 검색의 대표적인 방법으로 GRNet, FLAM이 있다.

## III. The Proposed Scheme

본 연구는 환자복과 평상복 구분을 위해 ResNet 모델과 YOLO 모델을 구축하고 실시간으로 객체 인식 및 객체 구별 프로세스를 수행하도록 하여 연구가 목표로 하는 바를 이루고자 한다. 가장 먼저 ResNet 모델의 개념을 설명하고 이후 ResNet 모델의 단점을 보완하고 한계를 극복하고자 개발한 YOLO 모델의 개념을 설명한다. 또한 모델을 학습시킬 때 사용한 학습데이터 생성과 이 데이터를 바탕으로 모델을 학습시킨 과정을 순서대로 설명하며 최종적으로 연구 결과가 지닌 의의와 한계를 제시한다.

### 1. ResNet Model

ResNet Model이란 2015년 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)대회에서 우승을 차지한 깊은 계층 학습을 배경으로 한 CNN 기반 이미지 분류 모델이다. 마이크로소프트사에서 개발하였으며, 저널 논문인 "Deep Residual Learning for Image Recognition"에서 소개되었다[14].

Top-5 error는 ResNet152에서 3.57%를 기록하며 이전 모델들과 비교되는 월등히 좋아진 성능을 보여주었다. ResNet 모델은 발표 직후 유용성을 인정받아 이미지 분류 분야의 연구들과 각종 소프트웨어 개발에 사용되었다.

인공지능 신경망의 계층을 깊게 한다고 해서 신경망의 성능이 무조건 좋아지는 것은 아니다. Vanishing/Exploding Gradients 문제가 발생해 오히려 신경망의 성능을 저하시킬 수도 있다. 그러므로 ResNet은 신경망을 깊게 구축하면서 그 성능을 저하시키지 않도록 할 기술을 필요로 했다. 그 기술이 Fig. 1의 Residual Network이다.

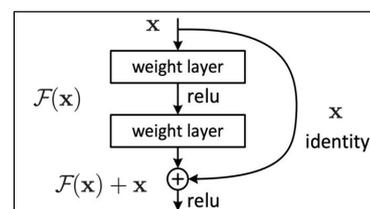


Fig. 1. Residual Network

Residual Network란 입력 값을 출력 값에 바로 더할 수 있도록 지름길(shortcut)을 추가해준 것을 말한다.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10 <sup>9</sup>	3.6×10 <sup>9</sup>	3.8×10 <sup>9</sup>	7.6×10 <sup>9</sup>	11.3×10 <sup>9</sup>

Fig. 2. ResNet Architectures

앞에서 ResNet은 그 유용성을 인정받아 많은 개발에 응용되었다고 소개했다. 이 유용성에는 모델의 큰 변화 없이도 코드를 조금만 수정해 쉽게 여러 응용에 접목시킬 수 있다는 강점이 포함된다. 본 연구에서는 이런 강점을 배경으로 기존의 ResNet의 정규모델인 Fig. 2와 ImageNet를 참고하여 신경망을 구축했고, 이후 여러 학습과정을 거치며 본 연구의 학습에 적합하도록 신경망을 다듬어 갔다. ResNet은 신경망이 깊어질수록 더 좋은 성능을 제공해주지만 본 연구에는 다음과 같은 문제점과 제약점 때문에 신경망을 무조건 깊게 할 수 없었다.

- 신경망이 깊어질수록 더 많은 연산을 수행해야 하므로 고성능의 CPU와 GPU가 필수불가결적으로 요구된다. ResNet-50의 파라미터 개수는 25.56M개이고, ResNet-152의 파라미터 개수는 60.19M개이다.
- 본 연구의 개발환경은 CCTV이다. 그러므로 제한된 성능 내에서 신경망이 구성되어야 하고, 또한 실시간 처리가 가능해야 한다.
- 속도 향상을 위해 수행되는 최적학습과 추가학습을 잘못 수행하게 된다면 높은 에러율을 발생시킬 수 있다.

그러므로 제한된 성능 내에서 최적의 에러율을 가지도록 성능과 사양 면에서 가장 적절한 절충안을 찾아야만 했다.

이런 절충안을 만족시킨 신경망의 개요는 다음과 같다. 신경망은 ResNet-50을 기초로 구축되었고, 학습은 Supervised Learning으로 수행되었다. 입력으로 사람이 입은 옷이 포함된 이미지 데이터를 받고, 해당 이미지가 환자복에 가까운지, 평상복에 가까운지 0~1 사이의 값으로 각각 점수를 매겨 높은 점수를 갖는 클래스로 그 결과를 출력한다. 처음 구축된 신경망의 초기 정확도는 94%였다. 여기서 정확도란 입력 데이터를 환자복인지 평상복인지 신경망을 통해 구별 후 그 결과를 출력하여 출력결과와 입력데이터가 일치하는 정도 또는 확률을 의미한다.

이후 신경망 구조 변경, learning rate value 최적화, training data 개선 등을 통해 최종 정확도는 96.50%(error 3.5%)로 향상시켰다.

## 2. YOLO Model

YOLO(You Only Look Once)란 Real-Time Object Detection을 지원하는 시스템이다. 즉, 객체의 종류와 위치를 추측하는 것을 도와주는 딥러닝 기반 물체인식 알고리즘이다. YOLO의 강점을 한 가지 뽑자면 알고리즘이 다른 Detection 알고리즘보다 간단해 처리과정이 단순하고 속도가 매우 빠르다는 점이다. 하지만 다른 알고리즘에 비해 상대적으로 낮은 정확도를 보여준다는 단점을 내포하고 있다. 그러나 실시간 처리 시스템에서는 계속되는 감시 속에서 한번의 Detection만 발생해도 정보를 획득할 수 있고 이 정보로 동작할 수 있기에 낮은 정확도는 크게 문제되지 않는다. ResNet은 YOLO에 비해 성능은 보장되지만 검출 후 결과 출력까지 8.95s(모델 호출-객체 평가)라는 많은 시간이 소요된다. 그러므로 실제 사용처에서 빠른 처리가 필요할 경우를 대비해 YOLO모델을 활용한 추가적인 시스템 구현을 고려해보았다.

Type	Filters	Size	Output
Convolutional	32	3 × 3	256 × 256
Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1
	Convolutional	64	3 × 3
Residual	128 × 128		
	Convolutional	128	3 × 3 / 2
	Convolutional	64	1 × 1
2×	Convolutional	128	3 × 3
	Residual	64 × 64	
Convolutional	256	3 × 3 / 2	32 × 32
	Convolutional	128	1 × 1
	Convolutional	256	3 × 3
8×	Residual	32 × 32	
	Convolutional	512	3 × 3 / 2
8×	Convolutional	256	1 × 1
	Convolutional	512	3 × 3
	Residual	16 × 16	
4×	Convolutional	1024	3 × 3 / 2
	Convolutional	512	1 × 1
	Convolutional	1024	3 × 3
	Residual	8 × 8	
Avgpool	Global		
Connected	1000		
Softmax			

Fig. 3. DarkNet-53 Architecture

본 연구에서 YOLO를 응용하여 환자복/평상복 구분 알고리즘을 구축한 방법은 다음과 같다. 모델은 YOLO에서 제공하는 DarkNet-53 모델을 사용하였다. DarkNet-53은 병목구조(1X1, 3X3)와 skip connection으로 구성되어 있다. 학습 시 입력데이터로 환자복과 평상복이 포함되고 Labeling이 완료된 정지된 이미지를 받으며 출력으로 98개 grid block의 정보(위치, 폭, 높이 등)와 환자복/평상복 분류 클래스 데이터를 출력한다. 실제 테스트에서는 학습된 YOLO모델에 동영상 데이터를 입력하고 출력으로 해당 동영상에 인식한 객체를 둘러싼 직사각형과 환자복/평상

복을 구별하는 텍스트 데이터를 덧씌워 출력한다. 환자복/평상복 Class 예측 수행 시 시그모이드를 활용한 2진 분류로 동작하도록 하였다.

### 3. Patient Uniform/Casual Clothing Recognition System

ResNet과 YOLO는 환자복/평상복 객체 검출을 수행하는데 서로 다른 강점을 보여준다. 하지만 둘의 강점을 모두 만족하는 시스템을 제한된 성능을 가지는 CCTV에서 구현하는 것은 매우 어렵다. 그러므로 해당 시스템이 동작되는 환경에 맞도록 강점을 선택할 수 있게 두 모델을 따로 구현하였다.

#### 3.1 ResNet

ResNet은 YOLO와 비교해볼 때 상대적으로 속도 면에서 많이 부족한 모습을 보여주지만 성능 면에서는 매우 우수한 결과를 보여준다.

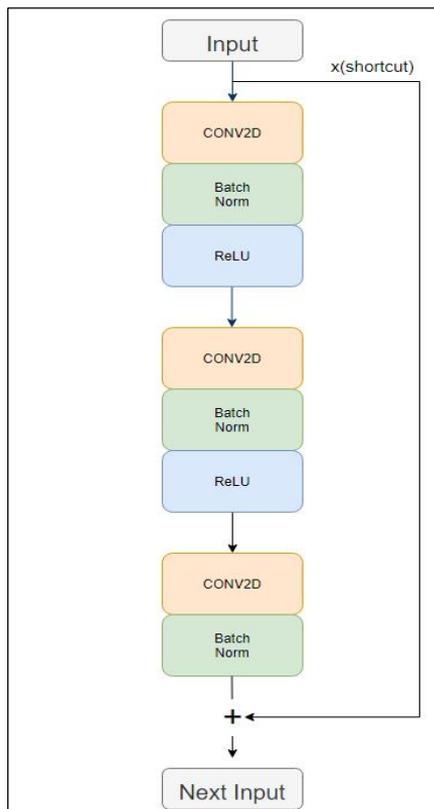


Fig. 4. ResNet 3-Layer

본 연구에서 ResNet모델은 50계층이며 Convolution연산과 Shortcut으로 구성되었다. 첫 계층과 마지막 계층을 제외하고 이전 계층으로부터 입력을 받으면 해당 입력을

$x$ (shortcut)로 저장한 뒤 학습을 수행한다. 여기서  $x$ 는 Residual Network를 구현할 때 사용되며 학습을 거치지 않은 이전 입력을 현재 출력으로 직접 연결한 데이터를 의미한다. 하나의 계층은 Convolution2D 연산, Batch Normalization 연산, Activation function ReLU 연산으로 구성된다. 각 Convolution은 서로 다른 필터의 개수와 크기를 가지며 구조는 Fig. 4를 반복하되 조금씩 변형시켰다.

#### 3.2 YOLO3

YOLO는 ResNet과 비교해볼 때 상대적으로 성능 면에서 많이 부족한 모습을 보여주지만 속도 면에서는 매우 우수한 결과를 보여준다.

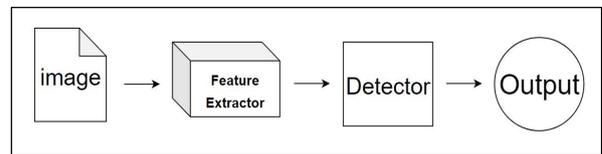


Fig. 5. YOLO Network

본 연구에서 구현된 YOLO 모델은 크게 4부분으로 구분될 수 있다.

1. image는 입력 데이터를 총칭하며 정지된 이미지, 동적인 동영상을 포함한다. 학습에는 정지된 이미지를 사용하고, 테스트와 실사용에는 동적인 동영상을 입력으로 받는다.
2. Feature Extractor는 Darknet-53을 의미한다. 입력 받은 이미지로부터 Multi-scale features를 추출하고 Detector의 3개의 입력으로 전달한다.
3. Feature Extractor로부터 특징정보를 받고 Bounding Box와 Class를 예측한다.
4. Output은 출력 데이터를 총칭한다. Bounding Box와 Class를 입력받은 영상 위에 덧씌워 출력한다.

## IV. Experimental Results

### 1. System Environment

환자복/평상복 인식 시스템 YOLO 모델은 Table 1과 같은 환경에서 구현되었다. ResNet-50 모델은 Google에서 제공하는 Colaboratory 환경에서 구현되었다.

Table 1. YOLO System Environment

Category	Specification
OS	MS Windows 10
CPU	Intel(R) Core(TM) i7-6700 3.4GHz
GPU	NVIDIA GeForce GTX 970
RAM	8.00GB
Development Tool	Pycharm Ver.2020.2
Language	Python 3.7
Software API	CUDA 9.0, cuDNN 7.6.4 YOLO3, DarkFlow, Binary 1.8.1

Table 2. ResNet-50 System Environment

Category	Specification
Development Tool	Google Colaboratory
Language	Python 3.6.9
Software API	Keras

## 2. Learning

### 2.1 ResNet

ResNet은 CNN을 기초로 하여 계층을 깊게 만들고, Residual Network를 추가한 모델이다. 그러므로 학습 과정도 CNN과 매우 유사하다. 본 연구의 ResNet 모델의 학습 과정은 다음과 같다.

#### 1. Collection and Classification of Learning Data

ResNet은 다른 신경망에 비해 상대적으로 깊은 신경망을 가지므로 신경망의 많은 파라미터를 학습시켜야 하는 특징을 가지고 있다. 그러므로 많은 학습데이터를 필요로 하는데 요양병원에서 제공받은 CCTV 영상에서 획득한 환자복 데이터만으로는 학습하기에 한 없이 부족하였다. 그래서 부족한 데이터를 보충하고자 Web Crawling기법을 사용하여 Google Web Site에서 기존의 CCTV에서 획득한 데이터와 유사한 의복 데이터를 추가적으로 수집하였다.

이렇게 수집한 데이터는 Google Drive에 test와 train폴더를 만들고 다시 해당 폴더 각각에서 평상복과 환자복 폴더를 만들어 각 폴더에 중복되지 않도록 저장하였다.



Fig. 6. Casual Clothing data



Fig. 7. Patient Uniform data

ResNet모델은 학습데이터가 많을수록 보다 나은 성능을 보여준다. 하지만 Web Crawling으로 품질 좋은 데이터를 모으기 위해서는 복잡한 선별작업을 필요로 한다. 그러므로 Image Generator방법을 사용해 Crawling으로 얻은 기존의 데이터를 조금씩 변형하여 학습 데이터에 추가해 사용하였다. Keras에서 제공하는 Image Data Generator API를 사용하였고, 기존의 이미지에 회전, 대칭 이동, 수평 및 수직 이동 등의 옵션 파라미터 값들을 랜덤으로 설정하여 변형된 추가 데이터를 확보하였다.



Fig. 8. Transformed data

Table 3. Total Number of Learning Data

Data Category	Count(Casual, Patient)
train	(250, 250)
test	(100, 100)
Generator_train	(750, 750)
Generator_test	(100, 100)

#### 2. Create Model and Invoke Learning Data

ResNet50은 Google Colab에서 '.ipynb' format으로 구축했다. 학습은 ResNet50이 구현된 .ipynb 파일에서 Google Drive 디렉토리를 호출하여 학습 데이터를 하나하나 읽어오며 학습시켰다.

#### 3. Learning and Save the Model

학습은 ResNet 모델의 구조, Learning rate, Total Number of Data 수를 조금씩 바꿔가며 여러 번 반

복하여 수행했고 가장 좋은 성능을 보여주는 모델은 load\_model API를 사용해 모델의 구성정보, 가중치 등의 파라미터를 수집하여 '.h5' format으로 저장했다. 모델을 저장해두면 필요할 때 학습된 결과를 호출할 수 있으므로 추후 모델 사용 시 실행 시간을 크게 단축시켜 준다.

## 2. 2 YOLO3

YOLO의 학습은 가장 먼저 데이터를 확보하고, 확보한 데이터의 Labeling 작업을 필요로 한다. 이후 Labeling 한 데이터를 사용하여 DarkNet을 학습시키고 학습시킨 모델에 테스트 데이터 또는 영상 데이터를 입력으로 넣고 학습 결과를 확인하는 순으로 진행된다.

### 1. Data Collection and Labeling

환자복과 일상복을 구별할 수 있도록 도와주는 공용 데이터는 존재하지 않아 직접 데이터를 수집하였다. 요양병원으로부터 CCTV녹화영상을 지원받았고, CCTV영상에서 사람이 입고 있는 의복을 캡처한 뒤, 캡처한 데이터를 Labeling하는 과정을 거쳐서 YOLO3에서 학습 가능한 데이터로 변환시켰다. Labeling Tool로 Binary 1.8.1을 사용하였다.

```
<?xml version="1.0"?>
<annotation>
  <folder>dataset</folder>
  <filename>81.png</filename>
  <path>C:\Users\Minji\Downloads\project\darkflow-master\generatordata\dataset\81.png</path>
  <source>
    <database>Unknown</database>
    <source>
  </source>
  <size>
    <width>1600</width>
    <height>900</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>doctor</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>752</xmin>
      <ymin>312</ymin>
      <xmax>1311</xmax>
      <ymax>784</ymax>
    </bndbox>
  </object>
</annotation>
```

Fig. 9. Labeling Data XML File

하나의 image에 대하여 Labeling을 수행하면 Fig. 9 와 같은 XML 파일을 획득 할 수 있다. 이 XML파일에는 image의 path 경로, 이미지의 가로X세로 크기, 관심 있는 영역(본 연구에서는 의류)의 bounding box와 같은 데이터를 저장하고 있다. 총 데이터 수는 image data 240개, 각 이미지에 1개씩 대응되는 XML file 총 240개이다.

### 2. Learning and Relearning

학습은 YOLO Mark Tool과 DarkFlow을 사용하여 진행하였다. 빠른 실시간 처리는 기존의 Yolo9000.cfg(상대적으로 느리나 정확도가 높음)에서 Tiny-YOLO.cfg(상대적으로 빠르나 정확도가 낮

음)로 모델을 변경하여 달성하고자 하였다. YOLO 모델을 학습 시킨 후 테스트 동영상을 모델에 입력으로 넣어주어 모델 학습 결과를 확인한다. 이 때 인식을 이 낮은 의류의 부위나 사람의 신체 부위가 발견되면 그 부위의 Labeling data를 추가하여 모델을 재학습 시키는 과정을 거친다. 이후에도 또 다시 결과를 확인하고 재학습 시키는 과정을 반복하며 모델의 정확도를 점점 높여주는 방식으로 학습을 진행하였다.

## 3. Experiment Results and Analysis

### 3. 1 ResNet

ResNet은 속도보다는 정확도에 초점을 맞춰 모델을 구현하고자 했다. 연구에 최적화된 모델을 찾아가는 과정에서 유의미 했던 결과에 대하여 정리하도록 하겠다. Table 4는 ResNet 모델 구현 전 개발한 CNN모델들이다. CNN Model 명은 ‘.’를 기준으로 왼쪽은 Loss Function을 오른 쪽은 (Convolution Stride, Pooling Stride)를 의미한다.

Table 4. Loss Function, Stride Change Result

CNN Model	Accuracy(%)
SoftMax-Cross Entropy, (2, 2)	60
SoftMax-Cross Entropy, (4, 2)	80
Mean Squared Error, (2, 2)	30
Mean Squared Error, (2, 2)	40

CNN모델은 생각보다 높은 성능을 보여주지 못했다. 이후 ResNet 모델개발로 연구방향을 전환했으며 파라미터 값들은 CNN모델에서 우수한 결과를 보여줬던 set를 가져와 약간 변형하여 사용하였다. ResNet Model 최종결과는 Table 5와 같다.

Table 5. ResNet Model Result

ResNet Model	Accuracy(%)
ResNet-50	94.00
ResNet-152	95.50
Learning Rate Value (ResNet-50)	
1.0X10 <sup>-5</sup>	92.50
5.0X10 <sup>-5</sup> (Default Value)	94.00
1.0X10 <sup>-4</sup>	96.50
3.0X10 <sup>-4</sup>	90.00
Total Number of Learning Data	
2000	94.00
3000	95.25
4000	95.50

ResNet152는 ResNet50보다 1.5% 더 높은 Accuracy를 갖는 결과를 보여줬지만 ResNet152는 ResNet50보다 시간 면에서 상대적으로 너무 오래 걸려 ResNet50모델을 최종

모델로 결정하였다. 이후 학습 상수를 여러 값으로 바꿔가며 적절한 값을 획득하고자 했고, 학습 데이터 개수와 학습 데이터들을 변경해가며 학습을 진행해 결과를 도출했다.

### 3. 2 YOLO3

YOLO3는 Darkflow를 활용하여 파라미터 값들을 조정 해주면 자동으로 학습이 이루어지는 매우 편리한 기능을 제공해준다. 또한 YOLO3는 실시간 처리를 제공해주기에 높은 정확도를 가지고 있지 않아도 한 번의 인식한 정보만을 가지고 결과를 도출할 수 있다는 장점을 가지고 있다. 그러므로 YOLO3는 성능보다는 속도에 중점을 두고 모델을 구현하였다. YOLO3 입력데이터와 출력결과 확인으로 2020년 3월에 방영된 대한민국 드라마 “슬기로운 의사생활”을 활용하였으며 YOLO 모델의 최종 Accuracy는 89%이다.

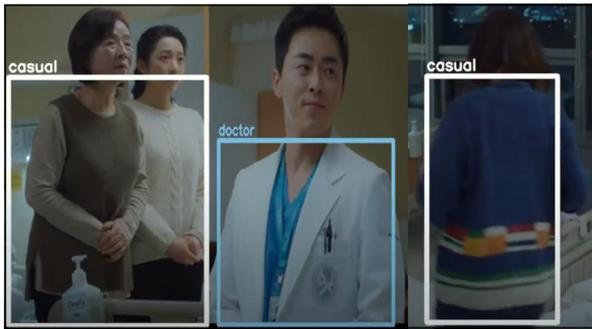


Fig. 10. Recognized Clothing

## V. Conclusions

본 연구의 목적은 노인인구 증가와 요양병원시설 등의 확대로 함께 증가하고 있는 노인안전사고 문제를 해결할 수 있는 방법을 제시해 노인안전사고 발생률을 감소시키는 것이다. 본 연구의 내용을 요약하자면 위험지역을 감지할 수 있는 위치에 CCTV를 설치하고, 해당 CCTV를 이용하여 위험지역에 접근하는 인물을 탐지, 의복을 인식한 후 환자복, 평상복을 구별하여 접근대상을 판별하도록 하는 인공지능 모델을 만드는 것이다.

본 연구를 진행하며 나온 결론으로 다음과 같은 의의와 판단을 내릴 수 있다. 첫째, 요양병원에서 어르신들의 안전미숙과 실무자들의 관리미흡 등으로 발생할 수 있는 안전사고를 줄일 수 있다. 둘째, 의료소비자들의 의식과 교육수준이 향상되면서 의료산업이 공급자 중심에서 수요자 중심으로 바뀌고 있는 현재 환자에게 더 나은 의료서비스

를 제공시켜주어 요양병원의 고객이탈을 막고, 소비자로 하여금 재이용 의도를 증진시킬 수 있을 것이다. 셋째, 물리보안 시장이 더욱 활성화되고 AI시장이 발전되고 있는 현재의 트렌드에 맞는 CCTV와 AI를 접목시킨 기술은 요양병원을 포함하여 더욱 많은 시설에 도움이 될 것으로 사료된다. 넷째, 본 논문의 의복인식 모델들은 성능이 좋고 널리 알려진 모델들을 기반으로 만들어져 의복의 형태와 색 뿐만이 아닌 섬유의 성질(면, 나일론 등)까지 고려하여 학습되어 더 높은 정확도를 획득할 수 있었다. 마지막으로 다섯째, 해당 기술은 약간의 변형만으로도 다른 환경 또는 다른 상황에 접목시킬 수 있다. 예를 들어 CCTV 감시 지역을 위험지역이 아닌 생활시설로 바꾸고, 구분할 대상을 환자복과 평상복이 아닌, 남 환자복, 여 환자복으로 바꾼다면(단, 남 환자복과 여 환자복의 패턴이 다르다는 가정 하에)남녀 환자의 남녀 병실 출입관리를 하는데 사용할 수 있을 것이다.

다만 본 연구는 다음과 같은 한계점을 가지고 있다. 첫째, 본 논문에서 개발된 것은 인식 시스템일 뿐 환자에게 이곳은 위험지역이라고 알려주는 기능은 존재하지 않는다. 인식 시스템만으로 노인의 안전사고 발생률을 감소시킬 수는 없을 것이다. 그러므로 최소한적으로 위험지역으로 접근하는 환자에게 접근하지 말라는 소리를 낼 수 있는 하드웨어의 도움을 필요로 할 것으로 보인다. 둘째, 대한민국에는 수 많은 요양병원이 존재하고, 그 요양병원마다 의복이 다양하므로 지금보다 더욱 포괄적이고 현실적인 데이터 수집이 필요하다. 또한 단순한 의복인식만으로 대상 판별을 하는 것은 한계가 있을 수 있다. 셋째, CCTV를 기준으로 사람이 다가오고 멀어짐에 따라 CCTV 화면에 들어오는 의상의 픽셀 수의 크기가 유동적으로 변하게 된다. 본 연구에서는 CCTV의 제한된 성능을 고려해 추가적인 많은 학습을 수행하지 못하여 CCTV에서 감지한 의복의 크기가 학습된 이미지의 크기랑 비교할 때 상대적으로 너무 작거나 너무 크면 인식이 떨어진다는 한계점을 지니고 있다. 마지막으로 코로나로 인해 병원에서 실제 데이터를 수집하는데 어려움이 많아 학습데이터 수가 많이 부족했다. 이는 실용성이 떨어진다는 문제를 야기할 수 있다. 그러므로 향후 연구에는 실제 데이터 수집에 더욱 더 힘을 써 모델의 질적인 면을 향상시키는데 박차를 가할 것이다.

## ACKNOWLEDGEMENT

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea. (Grant Number : HI18C1648010119)

## REFERENCES

- [1] C. H. Song, South Korea enters 'aged society' over 14% over 65 years of age, <https://www.donga.com/news/Society/article/all/20180828/91705793/1>.
- [2] M. H. Choi, Korea enters an aging society, changes in the menopausal market, <http://www.businesskorea.co.kr/news/articleView.html?idxno=40925>.
- [3] M. S. Yun. and J. Y. Moon, "Need to Analyze Demand Considering Active Areas and Living Population when Expanding Social Service Facilities Due to the Increase in the Elderly Population", Seoul Institute Policy Project Research Report, pp. 1-19, 2018.
- [4] D. Y. Park, Nursing hospitals and nursing homes increased by 15% compared to the previous year, a total of 6833 in 2017, <https://www.dailymedi.com/detail.php?number=835542&thread=22r14>.
- [5] D. H. Kim, "Trend of Increase in the Number of Patients and Days of Hospitalization Due to the Increase in the Number of Sickbeds in Nursing Hospitals", HIRA policy trends, Vol. 7, No. 3, pp. 53-61, 2013.
- [6] E. H. Ha. and G. S. Park, "Factors Influencing Cognitive Impairment in Elders with Dementia Living at Home", Journal of Basic Nursing, Vol. 18, No. 3, pp. 317-327, 2011.
- [7] M. S. Lee. and B. S. Lee, "A Conversation Analysis of Communication between Patients with Dementia and Their Professional Nurses", Journal of Korean Academy of Nursing, Vol. 36, No. 7, pp. 1,253-1,264, 2006.
- [8] U. K. Choi, "The Effects of Perceived Medical Service Quality on Patient Satisfaction, Hospital's Reputation and Loyalty", Journal of The Korea Society of Computer and Information, Vol. 25, No. 1, pp 177-185, 2020, DOI: <https://doi.org/10.9708/jksci.2020.25.01.177>.
- [9] J. M. Song and A. R. Park and S. B. Lee, "Trend of Technology in Video Surveillance System", Journal of The Korea Society of Computer and Information, Vol. 25 No. 6, pp. 57-64, 2020, DOI: <https://doi.org/10.9708/jksci.2020.25.06.057>.
- [10] H. Y. Kim, "A Tracking Algorithm to Certain People Using Recognition of Face and Cloth Color and Motion Analysis with Moving Energy in CCTV", Journal of Korean Society of Costume, Vol. 50, No. 2, pp. 81-95, 2000, DOI:10.3745/KIPSTB.2008.15-B.3.197.
- [11] H. J. Lee. and S. J. Lee and M. H. Choi and S. Y. Kim and I. Y. Moon, "Fashion Image Searching Website based on Deep Learning Image Classification", Journal of Practical Engineering Education, Vol. 11, No. 2, pp. 175-180, 2019, DOI: 10.14702/JPEE.2019.175.
- [12] Matiur Rahman Minar. and Thai Thanh Tuan. and H. J. Ahn, "An Improved VTON(Virtual-Try-On) Algorithm using a Pair of Cloth and Human Image", Journal of Korea Industrial Information Systems Research, Vol. 25, No. 2, pp. 11-18, 2020, DOI: <http://dx.doi.org/10.9723/jksiiis.2020.25.2.011>.
- [13] S. M. Lee and S. C. Oh and C. H. Jung and C. I. Kim, "Visual Fashion Analysis Using Deep Learning: A Survey", The Journal of Korean Institute of Communications and Information Science, Vol. 45, No. 7, pp. 1,174-1,182, 2020, DOI: <https://doi.org/10.7840/kics.2020.45.7.1174>.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition", In CVPR, 2015, DOI:10.1109/CVPR.2016.90.

## Authors



Ye-Chan Yun is currently an undergraduate student of the Department of IT Information Engineering at Jeonbuk National University. His research interests include Artificial Intelligence and Image Processing.



Young-Tae Kwak received the B.S., M.S., and Ph.D. degrees in computer engineering from the Chungnam National University, Republic of Korea, in 1993, 1995, and 2001, respectively.

He joined the faculty of the Jeonbuk National University in 2002. His research interests include Computer Vision and Neural Networks.