

## RGB-D 영상을 이용한 Fusion RetinaNet 기반 얼굴 검출 방법

남은정<sup>1</sup> · 남충현<sup>1</sup> · 장경식<sup>2\*</sup>

### Face Detection Method based Fusion RetinaNet using RGB-D Image

Eun-Jeong Nam<sup>1</sup> · Chung-Hyeon Nam<sup>1</sup> · Kyung-Sik Jang<sup>2\*</sup>

<sup>1</sup>Graduate Student, Department of Computer Engineering, Korea University of Technology and Education, Cheonan, 31253 Korea

<sup>2\*</sup>Professor, Department of Computer Engineering, Korea University of Technology and Education, Cheonan, 31253 Korea

#### 요 약

영상 내 사람의 얼굴을 검출하는 얼굴 검출 작업은 다양한 영상 처리 어플리케이션 내 전처리 또는 핵심 과정으로 사용되고 있다. 최근 딥러닝 기술의 발달로 높은 성능을 내고 있는 신경망 모델은 2차원 영상에 의존적이며, 카메라 품질이 떨어지거나, 얼굴의 초점을 제대로 잡지 못하는 등의 영상 내 노이즈가 발생할 경우, 제대로 얼굴을 검출하지 못할 수 있다. 본 논문에서는 2차원 영상의 의존성을 낮추기 위해 깊이 정보를 함께 사용하는 얼굴 검출 방법에 대해 제안한다. 제안하는 모델은 기존 공개된 얼굴 검출 데이터 셋을 이용하여 깊이 정보를 사전에 생성 및 전처리 과정을 거친 후 학습하였으며, 그 결과, 평균 정밀도 기준 FRN 모델은 89.16%로 87.95%의 성능을 보인 RetinaNet 모델보다 약 1.2% 정도의 성능이 향상되었음을 확인하였다.

#### ABSTRACT

The face detection task of detecting a person's face in an image is used as a preprocess or core process in various image processing-based applications. The neural network models, which have recently been performing well with the development of deep learning, are dependent on 2D images, so if noise occurs in the image, such as poor camera quality or poor focus of the face, the face may not be detected properly. In this paper, we propose a face detection method that uses depth information together to reduce the dependence of 2D images. The proposed model was trained after generating and preprocessing depth information in advance using face detection dataset, and as a result, it was confirmed that the FRN model was 89.16%, which was about 1.2% better than the RetinaNet model, which showed 87.95%.

**키워드** : 딥러닝, 객체 검출, 얼굴 검출, 깊이 정보

**Keywords** : Deep Learning, Object Detection, Face Detection, Depth Information

Received 28 November 2021, Revised 27 December 2021, Accepted 11 March 2022

\* Corresponding Author Kyung-Sik Jang(E-mail:ksjang@koreatech.ac.kr, Tel:+82-41-560-1352)

Professor, Department of Computer Engineering, Korea University of Technology and Education, Cheonan, 31253 Korea

Open Access <http://doi.org/10.6109/jkiice.2022.26.4.519>

print ISSN: 2234-4772 online ISSN: 2288-4165

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서론

영상 처리 분야 중 하나인 얼굴 검출(Face Detection) 작업은 영상에서 사람의 얼굴을 검출하는 작업으로 마스크 착용 여부와 같이 다양한 영상처리 어플리케이션 내 전처리 과정 또는 핵심 과정으로 사용되고 있다[1].

최근에는 얼굴 검출 작업을 수행하기 위해 입력 영상과 영상 내 얼굴 위치만을 이용하여 별도의 알고리즘 없이 합성곱 신경망(Convolutional Neural Network)과 같은 신경망 모델을 이용하여 높은 성능을 내고 있다[2-4].

하지만, 얼굴 검출 작업에서 높은 성능을 보인 신경망 모델들은 영상 내 삼원색으로 이루어진 픽셀 값들을 사용하기 때문에 입력 영상의 색 정보에 의존하여 얼굴 데이터를 학습한다. 이러한 의존성은 카메라의 영상 품질이 떨어지거나, 카메라가 흔들려 얼굴의 초점을 제대로 잡지 못하는 경우 등 입력 영상 내 노이즈가 발생하게 되었을 때, 얼굴을 정확하게 검출하지 못한다는 문제를 야기할 수 있다.

본 논문에서는 2차원 영상에 대한 의존성을 낮추기 위해 2차원 영상과 깊이 정보를 함께 사용한 얼굴 검출 방법을 제안한다. 자세히는 2차원 영상으로 깊이 정보를 만드는 방법, 깊이 정보의 범위 보정을 위한 Remapping 방법, 기존 신경망 모델들은 2차원 영상만을 입력으로 받기 때문에 깊이 정보를 함께 입력 받을 수 있는 FRN(Fusion Retina Network) 모델을 제안한다.

제안하는 모델은 기존 공개된 얼굴 검출 데이터 셋을 이용하여 깊이 정보를 사전에 생성한 후, 모델을 학습하여 평가하였다. 그 결과, 평균 정밀도 기준 FRN 모델은 89.16%로 87.95%의 성능을 보인 RetinaNet[2] 모델보다 약 1.2% 정도의 성능이 향상되었음을 확인하였다.

## II. 관련 연구

신경망 모델을 이용한 얼굴 검출 작업은 입력 영상 내 관심 영역의 중심점, 높이, 너비 값을 찾아내는 바운딩 박스 회귀(Bounding Box Regression) 작업과 찾아낸 관심 영역이 얼굴인지 아닌지 파악하는 분류(Classification) 작업으로 이루어져 있다.

최근 제안된 신경망 모델을 이용한 얼굴 검출 방법은 바운딩 박스 회귀 작업과 분류 작업을 동시에 수행하는

1단계 검출 방법과 바운딩 박스 회귀 작업을 수행한 후, 그 결과들에 대해 분류 작업을 순차적으로 수행하는 2 단계 검출 방법으로 나누어 연구가 진행되고 있다.

먼저 1단계 검출 방법을 사용하는 대표적인 신경망 모델은 YOLO(You Only Look Once) 모델[3], RetinaNet 모델[2]이 있다. YOLO 모델은 입력 영상을 사전 정의된 개수만큼 격자 영역을 분할하여, 각 영역에 대해 특정 개수만큼 얼굴이 있는지 없는지 예측한다. 이 모델은 모델의 깊이가 깊어질수록 특징 맵(Feature Map)의 가로, 세로의 크기는 줄이고 채널은 늘려가는 단계적 추출 방식을 통해 입력 영상에 대해 추상화 과정을 거치게 된다. 이때, 모델의 깊이가 깊어질수록 추상화된 특징 맵은 입력 영상 내 작은 얼굴들에 대한 정보가 손실되게 되어 제대로 예측하지 못한다는 문제점이 있다.

이러한 문제점을 완화하고자 RetinaNet 모델은 상향식(Bottom-Up) 모델과 하향식(Top-Down) 모델로 설계된 FPN(Feature Pyramid Network)[4] 모델을 이용하였다. 이 모델은 상향식 모델로부터 생성된 작아진 특징 맵을 하향식 모델의 입력으로 사용하여 작은 얼굴들도 잘 잡아낼 수 있다는 장점을 갖고 있다.

다음으로 2단계 검출 방법을 이용한 대표적인 신경망 모델은 R-CNN(Region-based Convolutional Neural Networks)[5], Fast R-CNN[6], Faster R-CNN[7] 모델 등이 있다.

먼저 R-CNN 모델은 영상에 대해 선택적 탐색(Selective Search) 알고리즘을 이용하여 관심 객체가 있을만한 후보 영역을 선정한 후, 후보 영역들에 대해 각각 독립적인 합성곱 신경망을 이용하여 특징 벡터를 생성한 후, 서포트 벡터 머신(Support Vector Machine) 알고리즘을 이용하여 각 후보 영역이 얼굴인지 아닌지를 예측한다.

하지만 이 모델은 각 후보 영역에 대한 특징 맵을 추출하기 위해 각각의 합성곱 신경망 모델을 사용해야하기 때문에 속도가 느리고, 많은 저장 공간이 요구된다는 문제점이 있다. 이러한 문제점을 개선하기 위해 제안된 Fast R-CNN은 선택적 탐색 알고리즘을 사용하는 것은 동일하지만 입력 영상에 대해 하나의 합성곱 신경망 모델을 이용하여 영상 전체에 대한 특징 벡터를 추출하여 선택적 탐색 알고리즘을 사용한다. 또한, Faster R-CNN 모델은 후보 관심 영역을 추출하기 위해 영역 제안 네트워크(Region Proposal Network)를 사용하였으며, 이 네트워크는 선택적 탐색 알고리즘의 낮은 속도를 보완함

으로써 상대적으로 더욱 빠른 추론 속도와 정확도를 얻을 수 있다는 장점이 있다.

### III. 깊이 정보를 이용한 얼굴 검출 방법

본 논문에서 제안하는 얼굴 검출 방법은 그림1과 같이 2차원 영상과 해당 영상에 대한 깊이 정보를 입력하는 과정과 FPN 모델을 이용하여 얼굴을 검출하는 과정으로 이루어져 있다.

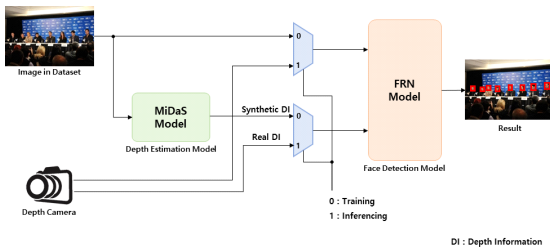


Fig. 1 Face Detection Diagram using Depth Information

입력 과정에서 사용되는 깊이 정보는 학습 과정과 추론 과정에 따라 다른 정보를 사용한다. 학습 과정에서는 사전 학습된 깊이 예측 모델인 MiDaS 모델[8]을 이용하여 가상의 깊이 정보를 생성한 후 사용하고, 추론 과정에서는 깊이 카메라를 이용한 실제 관측된 깊이 정보를 사용한다.

얼굴 검출에서 사용한 모델은 RetinaNet 모델을 기반으로 하며, 입력 단의 구조를 2차원 영상과 깊이 정보를 함께 입력 받을 수 있는 구조로 변형시킨 FPN 모델을 설계하여 사용하였다.

#### 3.1. 깊이 정보 생성 방법

공개된 얼굴 검출 데이터 셋들은 대부분 깊이 정보를 함께 포함하지 않기 때문에 본 논문에서는 사전 학습된 깊이 예측 모델인 MiDaS 모델을 이용하여 학습에 필요한 깊이 정보를 생성하였다.

기존 사전 학습된 깊이 예측 모델들은 학습, 평가를 위해 공개된 데이터 셋 중 각각 하나의 데이터 셋만을 사용하였다. 그 이유는 깊이 정보에 대한 실내, 외 촬영 환경, 측정 방법 등이 다르기 때문에 각각 다른 데이터 셋을 한꺼번에 다루기 어려웠기 때문이다. 하지만 MiDaS

모델은 공개된 다수의 데이터 셋을 사용하여 학습하였으며, 각각 다른 깊이 정보에 대한 손실을 보상하기 위해 학습 과정에서 손실 함수를 새롭게 정의하여 학습하였다.

그림2는 MiDaS 모델을 이용하여 얼굴 검출 데이터 셋에 대해 깊이 정보를 예측한 후 시각화한 결과이다. 이 그림을 통해 정확한 사람의 눈, 코, 입 등의 얼굴 내 특징들을 자세히 예측하지는 못하였지만, 사람의 얼굴, 몸에 대한 윤곽과 사람 간의 경계는 잘 구분하였음을 확인할 수 있었다.



Fig. 2 Visualization of Predicted Depth Information from RGB Images

#### 3.2. 깊이 정보 Remapping 방법

본 논문에서 사용하는 깊이 정보는 두 가지 문제점을 가지고 있다. 먼저, 먼 거리에 있는 사람들에 대한 깊이 정보가 불안정하다는 점이다. 그림2에서도 볼 수 있듯이 가까운 거리에 있는 사람의 윤곽이나 사람 간의 정보는 잘 표현되었음을 알 수 있지만, 먼 거리에 있는 사람들은 깊이 정보가 배경과 비슷하게 나타나 제대로 윤곽을 잡지 못하고 얼굴이 형체가 일그러지는 등 정보가 손실되는 경우가 발생한다.

또한 학습 과정에서 사용되는 가상의 깊이 정보와 실제 깊이 카메라를 이용하여 추론하는 과정에서 사용되는 실제 깊이 정보간의 인식 범위의 차이가 있다. 실제 깊이 카메라를 이용하여 촬영한 깊이 정보는 카메라의 깊이 측정 방식에 따라 달라질 수 있으며, 가상의 깊이 정보의 범위가 동일하지 않기 때문에 실제 깊이 정보를 이용하여 추론하는 경우 모델이 학습한 깊이 정보대로 입력이 되지 않을 수 있다.

이러한 문제점을 완화하기 위해 본 논문에서는 깊이 정보를 Remapping하는 작업을 수행하였으며, 이 작업은 표준화(Standardization), 정규화(Normalization), 로그 정규화(Log Normalization) 작업을 포함한다.

먼저 표준화 작업은 생성된 각 깊이 정보가 평균값을 기준으로 얼마나 떨어져있는지 상대적인 위치로 변환하기 위해 사용되었으며, 수식 (1)과 같다.

$$I_{i,j}^S = \frac{I_{i,j}^D - I_{mean}^D}{I_{std}^D} \quad (1)$$

좌표  $(i, j)$ 에 대해  $I_{i,j}^D$ 는 깊이 예측 모델이 예측한 깊이 정보이며,  $I_{mean}^D$ 은 깊이 정보의 평균값,  $I_{std}^D$ 는 깊이 정보의 표준편차,  $I_{i,j}^S$ 는 표준화된 깊이 정보 값이다.

다음으로 수식 (2)는 정규화 작업으로 서로 다른 피쳐들의 깊이 정보 값을 모두 0과 1사이의 상대 값으로 변경하기 위해 사용하였다.

$$I_{i,j}^N = \frac{I_{i,j}^D - I_{min}^D}{I_{max}^D - I_{min}^D} \quad (2)$$

정규화 작업 결과, 가장 깊은 깊이 정보는 1, 가장 가까운 깊이 정보는 0이 되며, 수식에서  $I_{max}^D$ 는 깊이 정보 중 가장 큰 값,  $I_{min}^D$ 은 가장 작은 값을 뜻한다.

마지막으로 로그 정규화 작업은 깊이 정보에 따라 객체들의 깊이 차이를 증폭시키기 위해 사용하였으며, 수식(3)과 같다.

$$I_{i,j}^{D'} = 1 - \log(I_{i,j}^D + 1) \quad (3)$$

### 3.3. Fusion Retina Network Model

기존 제안된 신경망 모델은 2차원 영상만을 입력으로 받아 얼굴을 검출하였기 때문에 깊이 정보를 함께 입력 받을 수 없다. 그리하여 본 논문에서는 그림3과 같이 기존 높은 성능을 보인 RetinaNet 모델을 기반으로 2차원 영상과 깊이 정보를 함께 입력 받을 수 있는 신경망 모델인 FRN(Fusion Retina Network) 모델을 설계하였다.

FRN 모델의 상향식 접근 방법은 기존 FPN 모델을 FFPN(Fusion Feature Pyramid Network) 모델로 변경하였으며, FFPN 모델은 독립적인 두 개의 동일한 신경망 모델로 이루어져있고, 입력으로 2차원 영상과 깊이 정

보가 각 신경망 모델에 입력된다.

각 단계별로 생성되는 특징 맵  $F_x$ 를 구하기 위해 입력 이미지의 특징 맵인  $C_x$ 와 깊이 정보 특징 맵인  $D_x$ 가 결합 연산을 하며, 가장 아래에 위치한 특징 맵  $F_1$ 는 큰 입력 데이터는 메모리를 많이 차지하기 때문에 제외하였다.

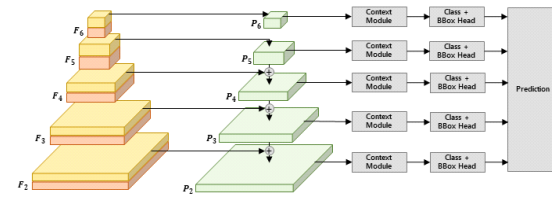


Fig. 3 Architecture of Fusion Retina Network Model

## IV. 실험 및 결과

### 4.1. 실험 환경

본 논문에서 제안한 모델의 학습 및 평가를 위해 사용한 얼굴 검출 데이터 셋은 WIDER 얼굴 데이터 셋[9]을 사용하였다. WIDER 얼굴 데이터는 32,203개의 영상과 393,703개의 얼굴에 대한 주석이 포함되어 있으며, 학습, 검증, 평가 데이터가 총 4:1:5 비율로 나누어져있다.

해당 데이터 셋에 대한 깊이 정보 데이터를 생성하기 위해 사전 학습된 깊이 예측 모델인 MiDaS 모델을 이용하였으며, WIDER 얼굴 데이터 내 12,881개의 학습 데이터와 3,226개의 검증 데이터에 대한 깊이 정보를 생성한 후, 학습 및 평가에 사용하였다. 학습은 총 30회 진행하였으며, Adam 최적화 알고리즘을 사용하였다. 또한, 신경망 모델의 가중치 업데이트를 위한 학습률 (Learning Rate)은 총 학습 횟수 30회에 대해 5단계로 변화하는 스케줄링 방법을 이용하여 학습을 진행하였으며, 이를 통해 학습이 진행됨에 따라 세밀한 가중치 조절이 가능하게 하였다. 먼저 학습이 시작되는 첫 번째 단계는 0.7로 설정하여 1회 학습하였으며, 두 번째 단계는 0.01로 설정하여 4회, 세 번째 단계는 0.05로 설정하여 5회, 네 번째 단계는 0.001로 설정하여 2회, 마지막 단계에서는 학습이 종료될 때까지는 0.0001로 설정하여 학습하였다. 이렇게 초기 학습 단계에 높은 학습률을 설정한 이유는 학습을 빠르게 진행하고자 하였기 때문이며, 0.7 이상의 경우 손실 값이 발산해버리는 경향이 있

었기 때문에 실험적으로 해당 값을 선정하였다.

Anchor Box의 파라미터 중 Scale 비율은 {1, 1, 2<sup>1/3</sup>, 2<sup>2/3</sup>}을 사용하였고 측면 비율은 1:1 비율만 사용하였다. 이렇게 생성된 Anchor Box의 개수는 FFPN 모델에서 생성된 각 특징 맵의 가로, 세로 크기에 따라 총 102,300개가 생성된다.

입력 데이터의 가로, 세로 크기는 동일하게 640으로 맞췄고, 크기가 맞지 않은 경우 모두 패딩을 적용하였다.

### 4.2. 성능 평가

표1은 RetinaNet 모델과 FPN 모델의 성능을 비교한 결과이며, 성능은 평균 재현율과 정밀도를 이용해 비교하였다. RetinaNet 모델은 RGB 영상만을 사용하여 학습하여 평가하였고, FPN 모델은 RGB 영상과 깊이 정보를 함께 사용하여 학습한 후 평가하였다.

**Table. 1** Average Recall and Precision about Each Model (IoU=0.5[3])

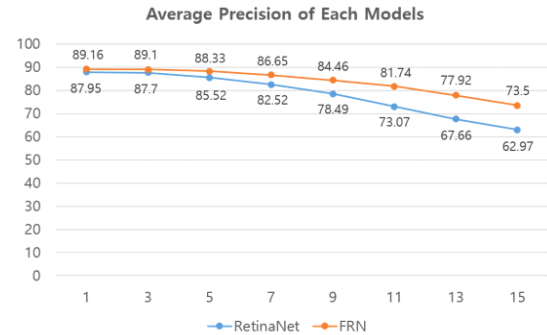
Models	Remapping Process	Recall(%)	Precision(%)
RetinaNet	not used	71.88	87.95
FPN	not used	73.82	87.83
	norm	74.64	88.97
	standard, norm	75.56	88.91
	norm, standard	71.13	88.22
	log, standard, norm	77.20	89.16

깊이 예측 모델로부터 예측된 깊이 정보에 대한 Remapping 처리는 norm, standard, log는 각각 표준화, 정규화, 로그 정규화를 의미하며, 왼쪽부터 순서대로 적용한 후 학습하였다. 예를 들어 log, standard, norm은 로그 정규화, 표준화, 정규화 작업을 순차적으로 한다는 뜻이다.

실험 결과, 실험의 베이스 모델인 RetinaNet 모델의 경우 정밀도 기준 약 87.95%의 성능을 보였으며, Remapping 처리를 모두 한 FPN 모델의 성능이 정밀도 기준 약 89.16%로 가장 높은 성능을 보였다. 또한 Remapping 처리 중 하나라도 처리한 경우 RetinaNet 모델 보다 약 1% 이상 높은 성능을 보였고, FPN 모델 또한 Remapping 처리를 하지 않은 경우 RetinaNet 모델과 유사한 정밀도 결과가 나왔음을 확인하였다.

### 4.3. 노이즈 영상에 대한 성능 평가

다음으로 낮은 품질의 카메라를 사용하는 경우 신경망 모델이 올바른 예측을 할 수 없게 될 때, 깊이 정보가 얼마나 성능을 보강해줄 수 있는지 노이즈 영상에 대한 성능을 평가하였다.



**Fig. 4** Average Precision about Image with Blue Noise

노이즈 영상은 본 논문에서 실험 데이터로 사용한 WIDER 데이터 셋에 노이즈를 주입하여 생성하였으며, 이때 인위적으로 주입한 노이즈는 블러링 노이즈이다. 블러링 노이즈는 크기가  $f$  인 합성 곱 필터를 이용하였으며, 총 8개의 필터 크기(1, 3, 5, 7, 9, 11, 13, 15)를 이용하였으며, 영상 내에 픽셀 한 칸씩 이동하면서 필터를 적용하였다. 여기서 필터의 크기가 1인 경우는 각 픽셀에 노이즈가 주입되기 때문에 합성 곱 필터가 적용되지 않은 것과 동일하다.

실험 결과, 그림4와 같이 필터의 크기가 커질수록, RetinaNet 모델의 성능이 약 24.98% 떨어진 것에 비해 FRN 모델은 약 15.66%의 성능만이 저하되었음을 확인할 수 있다.

그림5는 크기가 5인 필터를 적용한 블러링된 이미지와 원본 영상에 대해 RetinaNet 모델과 FPN 모델의 예측 결과를 시각화한 결과이다. 2차원 영상에 필터가 적용된 경우, 적용되지 않은 경우보다 예시에서 나타난 사람의 윤곽이 흐릿해지게 되며, 깊이 정보 또한 흐릿해진다. FRN 모델의 경우, 2차원 영상과 깊이 정보를 함께 사용하기 때문에 흐릿해진 영상과 깊이 정보가 상호보완을 해주어 영상 내 사람 5명의 얼굴을 모두 검출할 수 있었으며, 반면에 2차원 영상만을 이용한 RetinaNet 모델의 경우, 사람 5명의 얼굴 중 4명의 얼굴만을 검출하였음을 확인할 수 있었다. 이를 통해 2차원 영상과 깊이 정보를 함께 사용하는 경우 2차원 영상만을 사용한 모



텔보다 노이즈에 강건하다는 것을 실험을 통해 확인할 수 있었다.

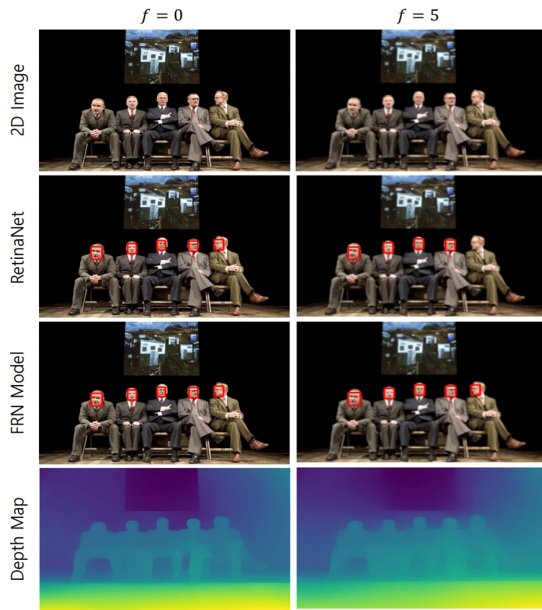


Fig. 5 Visualization of the Face Detected in the Original and Blurry Images

## V. 결론

얼굴 검출 작업은 영상에서 사람의 얼굴을 검출하는 작업으로 최근에는 얼굴 검출 작업을 수행하기 위해 입력 영상과 영상 내 얼굴 위치만을 이용하여 별도의 알고리즘 없이 합성곱 신경망과 같은 신경망 모델을 이용하여 높은 성능을 내고 있다.

하지만 기존 신경망 모델은 2차원 영상에 의존하기 때문에 카메라의 품질이 떨어지거나 노이즈가 발생한 경우 신경망 모델의 성능이 낮아지는 문제가 발생한다.

이러한 문제점을 완화하기 위해 본 논문에서는 2차원 영상과 깊이 정보를 함께 사용한 얼굴 검출 방법에 대해 제안하였다. 특히, 깊이 예측 모델을 이용하여 깊이 정보를 만드는 방법, 깊이 정보 범위 보정 방법, 깊이 정보 입력이 가능한 FRN 모델을 제안하였다.

제안하는 모델은 기존 공개된 얼굴 검출 데이터 셋을 이용하여 깊이 정보를 사전에 생성한 후, 모델을 학습하여 평가하였다. 그 결과, 평균 정밀도 기준 FRN 모델은

89.16%로 87.95%의 성능을 보인 RetinaNet 모델보다 약 1.2% 정도의 성능이 향상되었음을 확인하였다.

## ACKNOWLEDGEMENT

This results was supported by "Regional Innovation Strategy(RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE)(2021RIS-004)

## REFERENCES

- [ 1 ] C. H. Nam, E. J. Nam and K. S. Jang, "Mask Wearing Detection System using Deep Learning," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 25, no. 1, pp. 44-49, Jan. 2021.
- [ 2 ] T. Y. Lin, P. Goyal, R. Girshick, K. He and P. Dolla, "Focal Loss for Dense Object Detection," in *Proceeding of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, pp. 936-944, 2017.
- [ 3 ] Y. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once," in *Proceeding of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, pp. 779-788, 2016.
- [ 4 ] T. Y. Lin, P. Dolla, R. Girshick, K. He, B. Hariharan and Serge Belongie, "Feature Pyramid Networks for Object Detection," in *Proceeding of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, pp. 936-944, 2017.
- [ 5 ] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceeding of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Ohio, pp. 580-587, 2014.
- [ 6 ] Y. Redmon, S. Divvala, R. Girshick and A. Farhadi, "Fast R-CNN," in *Proceeding of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Las Condes, pp. 1440-1448, 2015.
- [ 7 ] S. Ren, H. He, R. Girshick, J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Proceeding of the 28th International Conference on Neural Information Processing Systems*, Montreal, pp. 91-99, 2015.

- [ 8 ] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler and V. Koltun, "Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer," arXiv 2019, arXiv:1907.01341[Online].
- [ 9 ] S. Yang, P. Luo, C. C. Loy and Xiaoou Tang, "WIDER FACE: A Face Detection Benchmark," in *Proceeding of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, pp. 5525-5533, 2016.



**남은정(Eun-Jeong Nam)**

2019년 부산외국어대학교 경제데이터금융학부 학사  
2021년 한국기술교육대학교 컴퓨터공학과 공학석사  
※관심분야 : Image Processing, Image Segmentation



**남충현(Chung-Hyeon Nam)**

2019년 한국기술교육대학교 컴퓨터공학부 공학사  
2019년 ~ 현재 한국기술교육대학교 컴퓨터공학과 공학석사과정  
※관심분야 : Natural Language Processing, Image Processing, Incremental Learning, Data Mining



**장경식(Kyung-Sik Jang)**

1987년 고려대학교 전자공학과 공학사  
1989년 한국과학기술연구원 전기전자공학과 공학석사  
1998년 동경공업대학 전기전자공학 공학박사  
1999년 ~ 현재 한국기술교육대학교 컴퓨터공학부 교수  
※관심분야 : Embedded System, Natural Language Processing, Incremental Learning