

# 초고속 R-CNN을 이용한 얼굴영상에서 눈 및 입술영역 검출방법

이정환  
안동대학교 전자공학과

## A Method of Eye and Lip Region Detection using Faster R-CNN in Face Image

Jeong-Hwan Lee  
Department of Electronic Engineering, Andong University

요 약 얼굴인식, 홍채인식과 같은 생체보안 분야에서 눈, 코, 입술 등 얼굴특징을 추출하는 과정은 필수적이다. 본 논문은 초고속(faster) R-CNN을 이용하여 얼굴영상에서 눈 및 입술영역을 검출하는 방법을 연구하였다. 초고속 R-CNN은 딥러닝을 이용한 물체검출 방법으로 기존의 특징기반 방법에 비해 성능이 우수한 것으로 알려져 있다. 본 논문에서는 얼굴영상에 콘볼루션, 선형정류과정, max pooling과정을 차례로 적용하여 특징맵을 추출하고 이로부터 제안영역(region proposal)을 검출하는 RPN(region proposal network)을 학습한다. 그리고 제안영역과 특징맵을 이용하여 눈 및 입술 검출기(detector)를 학습한다. 제안방법의 성능을 검토하기 위해 남녀한국인 얼굴영상 800장으로 실험하였다. 학습을 위해 480장을 이용했으며 테스트용으로 320장을 사용하였다. 컴퓨터모의 실험결과 눈 및 입술영역 검출의 평균정확도는 50 에포치일 때 각각 97.7%, 91.0%를 얻을 수 있었다.

주제어 : 딥러닝, 고속 R-CNN, 눈 및 입술영역 검출, 영상인식

**Abstract** In the field of biometric security such as face and iris recognition, it is essential to extract facial features such as eyes and lips. In this paper, we have studied a method of detecting eye and lip region in face image using faster R-CNN. The faster R-CNN is an object detection method using deep learning and is well known to have superior performance compared to the conventional feature-based method. In this paper, feature maps are extracted by applying convolution, linear rectification process, and max pooling process to facial images in order. The RPN(region proposal network) is learned using the feature map to detect the region proposal. Then, eye and lip detector are learned by using the region proposal and feature map. In order to examine the performance of the proposed method, we experimented with 800 face images of Korean men and women. We used 480 images for the learning phase and 320 images for the test one. Computer simulation showed that the average precision of eye and lip region detection for 50 epoch cases is 97.7% and 91.0%, respectively.

**Key Words** : Deep Learning, faster R-CNN, Eye and Lip Detection, Image Recognition

### 1. 서 론

최근 생체인식을 이용한 융합보안 분야가 주목받고

있으며, 얼굴인식은 생체인식분야에서 중요한 핵심연구 분야로 활발한 연구가 진행되고 있다[1-6]. 얼굴인식은 안면특징을 추출하여 각 개인을 인증하는 것으로 안면특

\*This work was supported by a Research Grant of Andong National University(2017).

\*Corresponding Author : Jeong-Hwan Lee (jhlee@anu.ac.kr)

Received March 2, 2018

Accepted August 20, 2018

Revised May 2, 2018

Published August 28, 2018

징인 눈, 코, 입, 귀, 얼굴형태 등의 특징추출이 매우 중요하다[3,5,6]. 특히 카메라로 획득된 영상에서 얼굴부분을 추출하기 위해 눈 및 입술영역을 찾아서 이를 기준으로 얼굴영역의 다른 특징 및 특징사이의 거리 등을 구하여 사용한다[3]. 특히 홍채인식을 이용한 정보보안 분야, 안구의 움직임 이용한 화면스크롤 등 시스템제어분야, 자동차 운전시 사람의 눈 움직임을 검출하여 사고를 방지하기 위한 졸음운전 예방시스템, 그 외 눈동자의 움직임을 이용한 뇌졸중 판단을 할 수 있는 헬스케어 등 다양한 응용분야에서 눈 및 입술영역 검출을 위한 활발한 연구가 진행되고 있다[5-7]. 따라서 생체인식분야에서 눈 및 입술영역을 검출하는 것은 매우 중요한 연구 과제이다.

본 논문에서는 초고속(faster) R-CNN(region convolutional neural network)을 이용하여 얼굴영상에서 눈 및 입술영역을 검출하는 방법을 연구한다. 기존 검출방법은 크게 특징기반 방법과 콘볼루션신경망 기반 방법으로 나눌 수 있다[3,5]. 특징기반 방법으로는 Harr-like 특징과 AdaBoost, SVM(support vector machine)알고리즘을 사용하여 눈 영역을 검출하는 방법이 있다[8]. 또한 색정보를 이용하여 눈 영역의 특징벡터를 만들고 이를 이용하여 눈 영역을 추출하는 방법이 있다[2-4]. 그러나 이들 대부분의 방법은 조명변화, 인경착용 여부, 얼굴의 회전 등에 따라 성능의 일관성이 부족하며 사용되는 특징에 의존적인 경향이 있다[5-8]. 이들 단점을 해결하기 위해 콘볼루션신경망 기반의 연구가 활발하게 진행되고 있다[10,19]. 콘볼루션신경망 기반은 딥러닝을 이용하는 것으로 입력영상으로부터 눈 및 입술영역에 해당하는 기준영역과 검출된 출력영역의 차이가 최소가 되게 콘볼루션의 필터계수를 신경망으로 학습하는 방법이다[9-11,18]. 이 방법은 학습데이터가 많을 경우 비교적 일관성 있는 성능을 유지하며 조명변화, 잡음첨가 등에도 강인한 특성을 나타낸다.

본 논문에서는 최근 다중물체 인식 알고리즘으로 활발한 연구가 진행되고 있는 초고속 R-CNN[11-13,17]을 이용하여 얼굴영상에서 눈 및 입술영역을 동시에 검출하는 방법을 제안하고 성능을 평가한다.

## 2. 초고속(faster) R-CNN 방법

본 절에서는 초고속 R-CNN의 근간이 되는 R-CNN[11],

고속(fast) R-CNN[12], 초고속(faster) R-CNN[13,17]에 대하여 설명한다. R-CNN은 콘볼루션 신경망을 이용하는 것으로 입력영상으로부터 제안영역(region proposal)을 사전에 추출하여 이를 CNN입력으로 사용하는데 CNN입력은 영상크기가 일정해야 하므로 제안영역의 크기를 확대 혹은 축소하여 사용한다. 여기서 제안영역이란 콘볼루션으로 구한 특징맵(feature map)에서 검출할 물체를 포함하고 있는 후보영역이다. 그리고 각 제안영역에 대하여 CNN을 독립적으로 적용하므로 제안영역이 서로 중복되는 경우도 많기 때문에 물체를 검출하는데 많은 시간이 소요되는 단점이 있다. CNN은 미리 학습된(pre-trained) AlexNet, VGG 등을 사용해서 특징을 추출하고, 분류기로 성능이 우수한 것으로 알려진 SVM을 특징맵의 가장 마지막 층에 적용하여 물체를 분류(classification)하는 방법이다[9,10,14]. 그러나 R-CNN은 모든 제안영역 마다 CNN을 별도로 실행해야하고 물체 분류기로 작동하는 SVM이 CNN과 통합되지 않아 분류기 결과를 CNN을 학습하는데 사용할 수 없는 단점이 있다. 또한 검출된 물체에 가장 잘 맞는 사각형박스를 결정하는 선형회귀(linear regression)모듈을 전체 시스템에 통합하여 학습시킬 수 없는 문제도 있다[11].

이러한 R-CNN을 개선한 것이 고속(fast) R-CNN이다. 즉 고속 R-CNN은 제안영역을 사전에 별도과정에서 독립적으로 추출하는 것은 R-CNN과 동일하지만, 검출된 각 제안영역마다 별도로 CNN을 실행하던 것을 하나의 CNN을 이용하여 특징맵을 추출하는 구조로 변경하여 시간을 대폭 단축하였다[12]. 입력영상에 CNN을 한번만 실행하여 특징맵을 생성하고 제안영역 정보를 특징맵의 마지막 층에 적용하여 관심영역(RoI: Region of Interest) 중 하나를 선택하는 RoI pooling과정을 수행한다. 그리고 RoI특징벡터는 영역분류를 위한 softmax단계와 검출한 물체에 최적으로 외접한 사각형을 구하기 위한 사각선형회귀(bounding-box regression)과정을 거친다. 결국 고속 R-CNN은 R-CNN의 단점을 대부분 개선해서 처리시간을 단축했다. 즉 CNN을 이용한 특징맵 생성, 분류기 및 사각선형회귀과정을 하나의 신경망으로 통합하여 실행시간을 단축했지만, 제안영역을 추출하는 과정은 R-CNN과 마찬가지로 별도로 수행하는 문제는 여전히 가지고 있다.

이런 문제를 해결한 것이 초고속 R-CNN방법이다[13]. 초고속 R-CNN은 영상내의 물체검출 알고리즘으로

최근 활발한 연구가 진행되고 있는데 상대적으로 고속으로 물체를 검출할 수 있어 주목받고 있다. 즉 제안영역 추출방법을 전체신경망에 포함하여 처리시간 단축과 학습의 일관성을 갖도록 한 것이다. 제안영역(region proposal)은 약 2000개정도 생성되는데 많은 시간이 소요되는 과정이다. 제안영역은 선택탐색(selective search)[15] 혹은 에지박스(edgebox)방법[16] 등을 이용하여 검출하는데 CNN 전단계에서 별도로 검출하였으나 초고속 R-CNN에서는 이를 CNN에 포함함으로써 고속 처리가 가능하게 하였다. 즉 제안영역을 검출하는 네트워크를 RPN(region proposal network)이라고 하며 외부에 별도로 존재하던 과정을 초고속 R-CNN에서는 CNN 내부에 통합하여 CNN의 네트워크 변수(parameters)를 공유하게 함으로서 입력부터 출력까지 전 과정을 하나의 네트워크로 운용할 수 있게 한 것이다. Fig. 1에 초고속 R-CNN의 개념도를 표시하였다.

결론적으로 R-CNN, 고속 R-CNN 및 초고속 R-CNN은 공통적으로 모두 CNN을 사용한다. 다만 제안영역을 추출하는 방법과 그들을 분류하는 방법이 다를 뿐이다. R-CNN과 고속 R-CNN은 제안영역이 CNN의 입력이 되어야 하므로 제안영역을 추출하는 과정이 CNN을 수행하기 전에 별도로 추가되어야 한다. 고속 R-CNN은 이 제안영역을 추출하는 과정이 CNN과 비교해서 많은 시간이 소요되고 전체 시스템에서 병목현상을

발생시키는 요인이었으나 초고속 R-CNN에서는 CNN의 학습 및 예측과정에 이를 포함시켜 한 영상의 처리시간을 단축시켰다[13].

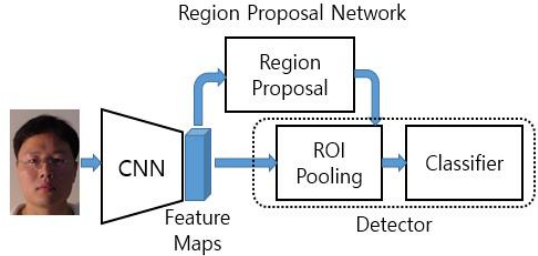


Fig. 1. Block diagram of faster R-CNN

### 3. 눈 및 입술 검출을 위한 제안방법

본 논문에서는 초고속 R-CNN을 이용하여 얼굴영상에서 눈 및 입술 위치를 검출하는 알고리즘을 연구하였다. 제안방법은 먼저 학습데이터의 입력과정, 초고속 R-CNN을 위한 CNN과정, 학습을 위한 변수설정 및 검출기 학습과정으로 구성된다. 제안방법의 블록도는 Fig. 2와 같다.

제안방법의 블록도를 살펴보면 알고리즘은 크게 RPN

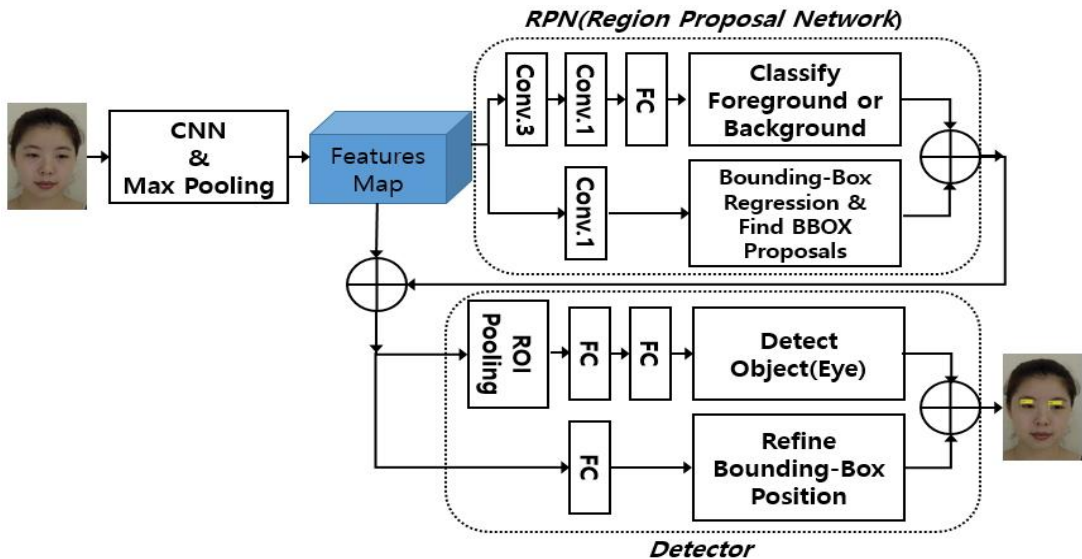


Fig. 2. Block diagram of the proposed method

과 검출기(detector)로 구성된다. 그리고 두 과정은 CNN 결과인 특징맵을 서로 공유하고 있음을 알 수 있다. 먼저 입력영상으로부터 CNN과 다운샘플링(max pooling)으로 특징맵을 추출한다. 이 때 CNN은 32개의 3x3크기 필터를 두 번 적용하고 마지막 층에서 2x2 max pooling으로 다운샘플링 한다. 본 논문에서 사용한 검출기의 특징맵의 크기는 14x19이며 채널수는 32개이다. 이 특징맵은 RPN의 입력이 되며 RPN은 다시 두 단계로 나누어진다. 첫 단계는 제안영역을 생성하는 부분으로 검출하려는 물체가 있는 전경(object) 영역과 물체가 없는 배경(not object) 영역을 구분하는 단계이다. 이 과정에서는 특징맵을 3x3 콘볼루션신경망(Conv.3)을 통과시키고, 다시 이를 1x1 콘볼루션신경망(Conv.1)을 통과시킨 후 이를 완전연결층(FC: fully connected layer)에 연결하여 배경영역과 전경영역을 분류한다.

그리고 RPN의 두 번째 단계는 전경영역의 위치와 전경영역에 외접하는 사각형(bounding-box)을 구하는 과정이다. 여기서 위치는 전경영역을 둘러싸는 사각형의 중심좌표( $x, y$ )이고 사각형의 크기는 가로 폭( $w$ )과 세로 높이( $h$ )로  $(x, y, w, h)$ 로 표시된다. 그리고 이 두 단계의 결과는 (전경, 배경)과  $(x, y, w, h)$ 의 6개 값이 되며 이를 이용하여 제안영역을 표현한다. 여기서 (전경, 배경)은 각각 0과 1사이의 실수 값을 갖는다. 이 값은 지수(anchor)박스와 기준(ground-truth)박스를 이용하여 Fig. 3과 같이 구한다. 지수박스는 눈 혹은 입술영역을 탐색하는 윈도우이다. 전경 및 배경을 구별하기 위한 값은 아래 식(1)을 이용한다[11,15,16].

$$P = \frac{A \cap B}{A \cup B} = \frac{S}{A \cup B} \quad (1)$$

여기서 A는 지수박스면적이고 B는 기준박스면적이다. S는 두 박스가 겹치는 면적이다. 본 논문에서는 전경은  $P \geq 0.5$  이상인 영역을 선택하고, 배경은  $P \leq 0.3$ 인 영역을 선택한다.

Fig. 2의 아래 부분은 검출기로 RPN과 마찬가지로 물체 검출과정과 눈 혹은 입술 영역을 둘러싸는 사각형을 물체에 가능한 외접하도록 미세 조정하는 두개의 과정으로 이루어진다. 먼저 물체검출과정은 ROI(region of interesting) pooling층, 두 개의 완전연결(fully connected layer)층, 눈 혹은 입술 검출과정으로 구성된다. 그리고

ROI pooling층의 출력은 두 개의 완전연결층(FC)을 거치는 및 입술영역과 배경으로 구분된다. 그리고 물체영역을 둘러싸는 사각형의 위치를 미세조정(fine-tune)하는 과정은 검출된 영역을 확정하기 위한 것으로 검출된 물체에 외접하는 사각형을 구하는 과정이다. 그 과정은 Fig. 4와 같다.

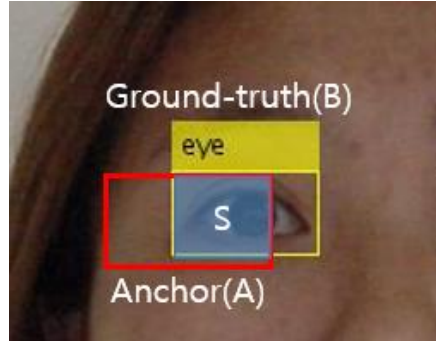


Fig. 3. Ground-truth and anchor box

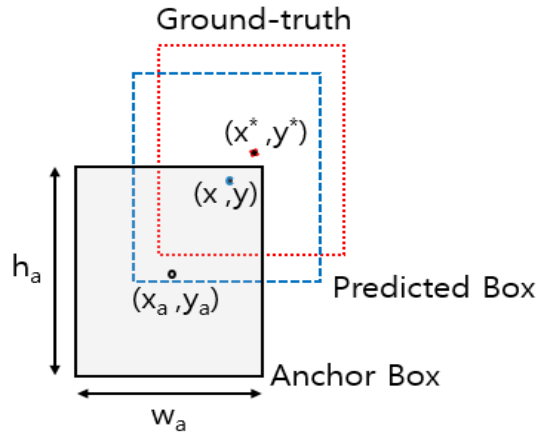


Fig. 4. Adjust an anchor box to ground-truth

여기서  $(x_a, y_a, w_a, h_a)$ 는 지수박스의 중심좌표, 폭, 높이를 의미하고,  $(x, y, w, h)$ 는 기준박스인  $(x^*, y^*, w^*, h^*)$ 에 근접하도록 반복하여 조정과정 중인 예측박스(predicted box)의 해당되는 데이터이다.  $(\Delta_x, \Delta_y, \Delta_w, \Delta_h)$ 는 예측박스가 기준박스와 일치하게 되도록 조정해야할 변환값으로 특징맵을 이용하여 구한다[13].

$$\begin{aligned}
 x &= w_a \Delta_x + x_a \\
 y &= h_a \Delta_y + y_a \\
 w &= w_a e^{\Delta_w} \\
 h &= h_a e^{\Delta_h}
 \end{aligned}
 \quad (2)$$

초고속 R-CNN의 학습은 4단계를 거쳐 이루어진다. 첫 번째 단계는 학습데이터를 이용하여 RPN을 학습한다. 두 번째 단계는 학습된 RPN으로부터 얻은 제안영역을 이용하여 검출기를 학습한다. 세 번째 단계는 블록도의 컨볼루션(CNN) 및 max pooling 단계의 가중치를 고정된 후 RPN을 미세조정 한다. 그리고 마지막 네 번째 단계는 세 번째 단계로부터 구한 제안영역을 이용하여 검출기를 미세조정한다. 여기서 세 번째 및 네 번째 단계에서는 컨볼루션 및 max pooling과정의 가중치를 서로 공유하게 된다. 또한 Fig. 2 블록도의 RPN을 제외하면 아래 부분의 검출기는 결국 고속(fast) R-CNN의 동작과 같다. 즉 외부에 별도로 존재하던 RPN을 네트워크 내부로 통합함으로써 블록도 앞쪽의 컨볼루션 및 max pooling부분을 공유하게 되어 물체 검출 시간을 단축할 수 있게 된다.

#### 4. 실험결과 및 고찰

제안방법의 성능을 평가하기 위해 본 연구에서는 남녀한국인 800장의 칼라 얼굴영상을 이용하여 눈 및 입술영역을 검출하는 컴퓨터모의실험을 한다. 초고속 R-CNN 학습용으로 480장을 이용하고, 테스트용으로 320장을 사용하였다. 실험은 Intel Xeon 2.2GHz CPU, 32GB 주기억장치, Titan X GPU(12GB) 및 64비트 운영체제에서 수행하였다.

Fig. 5와 Fig. 6에 눈과 입술 샘플영상을 나타내었다. 실험에 사용된 영상은 남녀 얼굴을 일정한 거리에서 상하좌우로 조금씩 회전하고 조명을 좌우에서 비추어 조명 변화가 발생하도록 하여 영상을 획득한 것이다. Table 1에 얼굴영상에서 초고속 R-CNN을 학습하기 위한 기준 영상인 눈 및 입술영역의 크기정보(최소, 최대, 평균)를 최소단위로 표시하였다.

Fig. 2의 블록도에서 첫 번째 단계인 CNN 및 max pooling과정은 컨볼루션층과 다운샘플링과정인 max pooling, 그리고 선형정류(ReLU: rectified linear unit)과정을 포함하는 것으로 RPN과 검출기에서 사용할 특징맵

을 구하는 과정이다. Fig. 7은 이 단계의 첫 컨볼루션층에서 칼라영상 (r, g, b) 채널에 순서대로 적용된 필터계수를 흑백영상으로 표현한 것이다. 필터크기 3x3, 채널수 3, 필터갯수 32개로 구성되며 정규화된 필터계수는 모두 96개이다. 필터계수값의 변화가 다양하고 이를 얼굴영상에 적용하면 여러 가지 형상의 특징을 추출할 수 있게 된다.



Fig. 5. Example of eye(positive sample) images

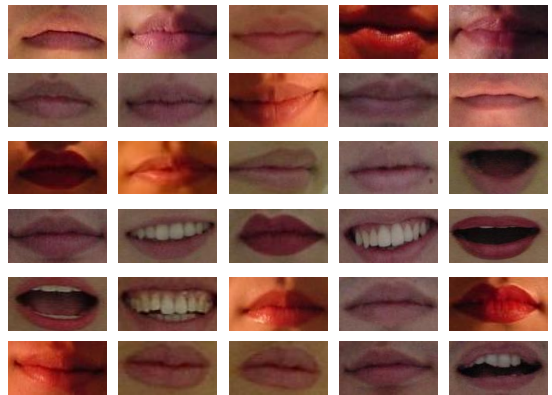


Fig. 6. Example of lip(positive sample) images

Table 1. Size of input image, eye and lip area for ground-truth(unit: pixel)

	information	width	height
Input image	min	230	333
	max	595	700
	average	411	530
eye area	min	33	16
	max	114	57
	average	65	33
lip area	min	50	24
	max	174	116
	average	99	48

큰볼루션을 적용할 때 각 필터의 상하좌우 이동 (stride)은 각각 1화소씩 움직이고, 영상 크기를 입력영상과 동일하게 유지하도록 지주박스 가장자리에 추가하는 padding값은 1로 하였다.

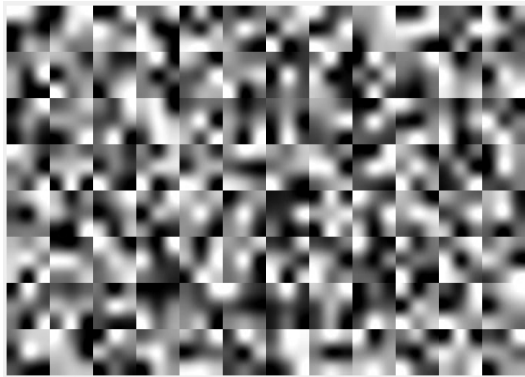


Fig. 7. Filter coefficient image for second convolution layer, which is 32 filters with 3x3

Fig. 8에 제안방법으로 눈과 입술을 검출한 예를 보였 다. Fig. 8 왼쪽은 눈 영역(노란색)만을 검출한 것이고, 오른쪽은 입술영역(녹색)만을 검출한 것이다. 여기서 파란 색은 눈의 기준영역이고, 빨간색은 입술영역의 기준영역이다. 그리고 Fig. 9는 눈과 입술을 동시에 검출한 결과이다. 검출된 결과는 기준영상과 겹치면서 얼굴에서 눈과 입술을 검출함을 보여준다. Fig. 10은 에러가 발생된 영상의 예를 나타낸 것인데, 조명, 눈 영역의 크기가 작으므로 눈 영역 검출시 입술영역 검출보다 상대적으로 에러가 발생한다. Fig. 11은 눈과 입술의 상대적인 위치를 역삼각형으로 표시한 것이다. 눈 및 입술위치는 검출된 사각형에서 대각선이 서로 만나는 좌표이다. 인식기술의 성능측정을 위해서 재현율(recall)과 정확도(precision)를 주로 사용하는데, 재현율은 대상물체를 빠뜨리지 않고 얼마나 잘 포착하는가를 나타내고, 정확도는 포착된 결과가 얼마나 정확한가를 의미하는 지표이다. Fig.12에 검출결과 중 일부를 표시하였다. Fig. 13은 50 epoch일 때의 눈 및 입술영역 검출의 평균정확도를 나타낸 것이다. 본 연구의 눈 및 입술영역 검출을 위한 평균정확도는 각각 97.7%, 91.0%이다. 웃는 영상에서 입술영역은 검출하지 못하는 경우가 일부 발생한다.

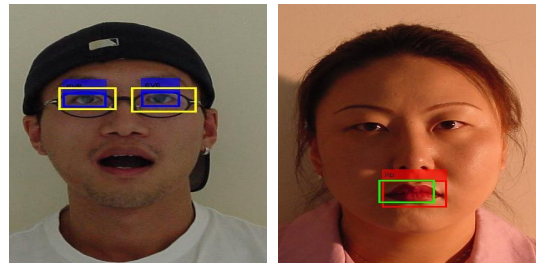


Fig. 8. An example of ground-truth(eye: blue, lip: red), separately detected eye(yellow) and lip(green), respectively

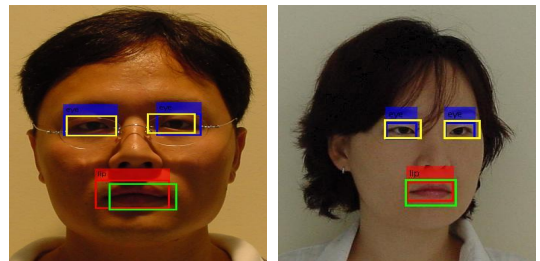


Fig. 9. An example of ground-truth(blue, red), simultaneously detected eye(yellow) and lip(green), respectively

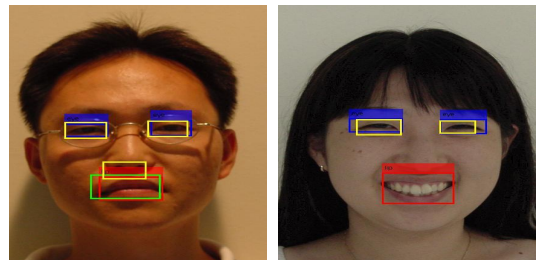


Fig. 10. An example of error occurred

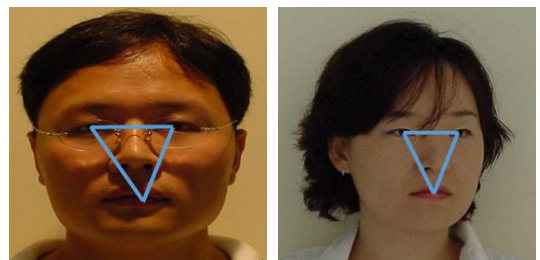
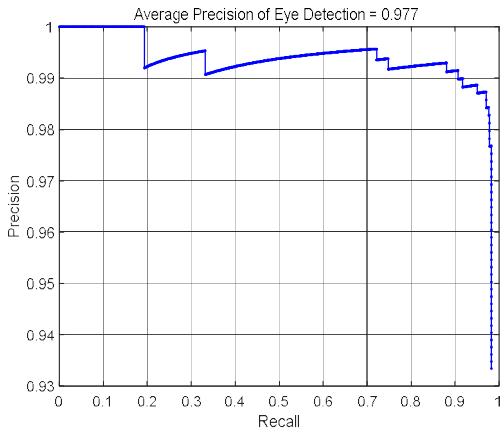


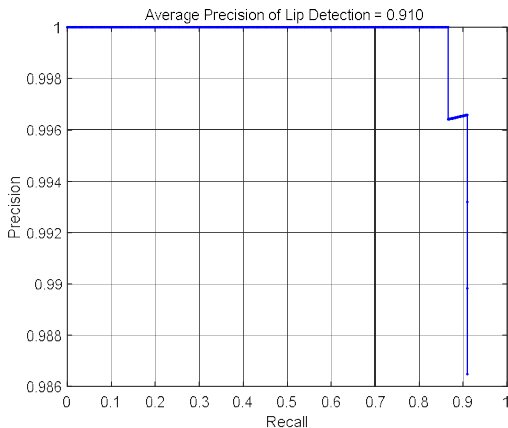
Fig. 11. Eye and lip position represented with an inverted triangle



Fig. 12. Samples of detection results



(a) average precision of eye detection



(b) average precision of lip detection

Fig. 13. Average precision of eye and lip detection in case of 50 epoch

## 5. 결론

본 논문에서는 초고속 R-CNN을 이용하여 얼굴영상에서 눈 및 입술 영역을 동시에 검출하는 방법을 연구하였다. 초고속 R-CNN은 고속 R-CNN에서 제안영역을 생성하는 과정을 신경망내부에 포함시켜 입력영상으로부터 물체 영역을 검출하는 전체 과정을 하나의 CNN에 통합한 것이다. 본 논문에서는 입력영상으로 부터 특징맵을 구하는 CNN을 연구하였으며, 제안영역을 생성하는 RPN과 눈 및 입술 영역을 검출하는 검출기를 실험하였다.

실험에 사용된 영상은 얼굴영상으로 부터 눈 및 입술 영역을 추출한 영상을 사용하여 RPN 및 검출기를 학습하였다. 컴퓨터모의 실험결과 제안방법을 얼굴인식을 위한 전단계로 눈 및 입술 영역을 검출하는데 사용될 수 있음을 알 수 있었으나, 앞으로 개선할 점들도 많이 있다. 평균정확도는 눈 영역이 입술영역 보다 상대적으로 높았는데, 이는 옷은 얼굴영상에서 이가 보이는 입술영역의 학습데이터 부족으로 입술영역을 검출하지 못하는 경우가 일부 발생한 것으로 향후 연구에서 개선이 필요하다. 그리고 얼굴영상에서 눈 영역뿐만 아니라 코, 귀, 턱, 머리부분 등 얼굴인식을 위한 주요특징을 검출하는 다중물체 검출시스템으로 확장하는 방법을 연구할 필요가 있다.

## REFERENCES

- [1] S. Zafeiriou, C. Zhang & Z. Zhang. (2015). A survey on face detection in the wild: past, present and future. *Computer Vision and Image Understanding*, 138, 1-24. DOI : 10.1016/j.cviu.2015.03.015
- [2] Vinay Kumar, Arpit Agarwal & Kanika Mittal. (2011). *Tutorial: introduction to emotion recognition for digital images*, [Technical report] <inria-00561918>, 1-47.
- [3] Oya Çeliktutan, Sezer Ulukaya & Bülent Sankur. (2013). A comparative study of face landmarking techniques. *EURASIP Journal on Image and Video Processing*, 13. DOI : 10.1186/1687-5281-2013-13
- [4] S. Y. Kwon et al. (2012). Comparative performance evaluations of eye detection algorithm. *Journal of Korea Multimedia Society*, 15(6), 722-730.
- [5] Dan Witzner Hansen & Qiang Ji. (2010). In the eye of the beholder: a survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine*

- Intelligence*, 32(3), 478 - 500.  
DOI : 10.1109/TPAMI.2009.30
- [6] Waqas, Haider, Hadia Bashir, Abida Sharif, Irfan Sharif & Abdul Wahab Smola. (2014). A survey on face detection and recognition approaches. *Research Journal of Recent Sciences*, 3(4), 56-62.
- [7] A. Al-Rahayfeh & M. Faezipour. (2013). Eye tracking and head movement detection: a state-of-art survey. *IEEE Journal of Translational Engineering in Health and Medicine*, 1.  
DOI: 10.1109/JTEHMH.2013.2289879
- [8] P. Viola & M. Jones. (2004). Robust real-time object detection. *International Journal of Computer Vision* 57(2), 137 - 154.  
DOI : 10.1023/B:VISI.0000013087.49260.fb
- [9] Y. L. Cun, Y. Bengio, & G. Hinton. (2015. May). Deep learning. *Nature*, 521(7553), 436-444.  
DOI : 10.1038/nature14539
- [10] J. Schmidhuber. (2015). *Deep learning in neural networks: an overview*. *Neural Networks*, 61, 85-117.  
DOI : 10.1016/j.neunet.2014.09.003
- [11] R. Girshick, J. Donahue, T. Darrell & J. Malik. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5.
- [12] Girshick, Ross. (2015). Fast r-cnn. *Proceedings of the IEEE International Conference on Computer Vision (ICCV 2015)*.
- [13] S. Ren, K. He, R. Girshick, & J. Sun. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137-1149.  
DOI: 10.1109/TPAMI.2016.2577031
- [14] A.J. Smola. & B. Schölkopf. (2004). A tutorial on support vector regression. *Statistics and Computing* 14(3), 199 - 222.
- [15] J. R. Uijlings, K. E. van de Sande, T. Gevers & A. W. Smeulders. (2013). Selective search for object recognition. *International Journal of Computer Vision*, 154-171.
- [16] C. L. Zitnick & P. Dollar. (2014). Edge boxes: Locating object proposals from edges. in *European Conference on Computer Vision (ECCV)*, 391-405.
- [17] <https://kr.mathworks.com/help/vision/examples>
- [18] Y. Zhang, K. Lee & H. Lee. (2016). Augmenting supervised neural networks with unsupervised objectives for large-scale image classification. *International Conference on Machine Learning (ICML)*, 612-621.
- [19] K. Kim, et al. (2017). Detail focused image classifier model for traditional images. *Journal of the Korea Convergence Society*, 8(12), 85-92.  
DOI: 10.15207/JKCS.2017.8.12.085

이 정 환(Lee, Jeong Hwan)

[정회원]



- 1983년 2월 : 경북대학교 전자공학과(공학사)
- 1993년 8월 : KAIST 전기및전자공학과(공학 석·박사)
- 1990년 3월 ~ 현재 : 안동대학교 전자공학과 교수
- 관심분야 : 인공지능, 영상처리, 생체인식, 워터마킹 등
- E-Mail : jhlee@anu.ac.kr