

# Gender Classification System Based on Deep Learning in Low Power Embedded Board

Hyunwook Jeong<sup>†</sup> · Dae Hoe Kim<sup>††</sup> · Wisam J. Baddar<sup>†††</sup> · Yong Man Ro<sup>††††</sup>

## ABSTRACT

While IoT (Internet of Things) industry has been spreading, it becomes very important for object to recognize user's information by itself without any control. Above all, gender (male, female) is dominant factor to analyze user's information on account of social and biological difference between male and female. However since each gender consists of diverse face feature, face-based gender classification research is still in challengeable research field. Also to apply gender classification system to IoT, size of device should be reduced and device should be operated with low power. Consequently, To port the function that can classify gender in real-world, this paper contributes two things. The first one is new gender classification algorithm based on deep learning and the second one is to implement real-time gender classification system in embedded board operated by low power. In our experiment, we measured frame per second for gender classification processing and power consumption in PC circumstance and mobile GPU circumstance. Therefore we verified that gender classification system based on deep learning works well with low power in mobile GPU circumstance comparing to in PC circumstance.

**Keywords :** Gender Classification, Deep Learning, Embedded Board, Low Power

# 저전력 임베디드 보드 환경에서의 딥 러닝 기반 성별인식 시스템 구현

정 현 옥<sup>†</sup> · 김 대 회<sup>††</sup> · Wisam J. Baddar<sup>†††</sup> · 노 용 만<sup>††††</sup>

## 요 약

사물 인터넷(IoT) 산업이 확산되면서 사용자의 정보를 특별한 조작 없이 물체가 스스로 인식하는 일이 매우 중요해졌다. 그중에서도 성별(남, 여)은 생물학적인 구조가 달라 성향이 다르고 사회적으로도 기대하는 바가 다르기 때문에 매우 중요한 요소이다. 하지만 얼굴 이미지를 기반으로 한 성별 인식과 관련된 연구는 동일한 성별이라도 다양한 생김새를 가지고 있어서 여전히 도전적인 분야이다. 그리고 성별인식 시스템을 사물 인터넷에 적용하기 위해서는 디바이스 크기를 소형화 시켜야 하며 저전력으로 구동이 가능해야 한다. 따라서 본 논문에서는 저전력으로 실제 사물에서 성별을 인식할 수 있는 기능을 탑재하기 위해 딥 러닝 기반의 성별 인식 알고리즘을 제안하고 이를 모바일 GPU 임베디드 보드에 포팅하여 최종적으로 실시간 성별인식 시스템을 구현하였다. 실험에서는 소비전력과 초당 처리 가능한 프레임 수를 PC환경과 모바일 GPU 임베디드 환경에서 측정하여 저전력 환경에서도 성별 인식이 가능함을 증명하였다.

**키워드 :** 성별 인식, 딥 러닝, 임베디드 보드, 저전력

## 1. 서 론

최근 사물인터넷(IoT) 기술이 산업계에서 뜨거운 이슈로

부각되고 있다. 사물인터넷은 저전력 환경에서 사물이 인터넷 네트워크로 연결되어 실제 환경에 부합되는 서비스를 인간에게 능동적으로 제공할 수 있는 기술을 뜻한다. 그러한 서비스를 제공하기 위해서는 사용자에 대한 사전 정보를 시스템이 능동적으로 인지할 수 있어야 한다.

사용자에 대한 사전 정보에는 성별, 연령, 감정, 사용자의 위치 등 여러 요소가 있다. 그 중에서도, 성별은 남, 여 각각에 대해 사회적으로 기대되는 역할이 다르고[1], 생물학적인 특성도 다르기 때문에 사용자의 정보를 파악하는데 있어서 중요한 역할을 한다. 하지만 같은 성별이라도 사람마다 얼굴 생김새나 머리 스타일 등이 다르기 때문에 성별인식은 지금까지 매우 도전적인 문제로 여겨졌다[1-2].

\* 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2015R1A2A2A01005724).

\*\* 이 논문은 2016년도 한국정보처리학회 춘계학술발표대회에서 '임베디드 환경에서의 딥 러닝(Deep Learning) 기반 실시간 성별 인식'의 제목으로 발표된 논문을 확장한 것임.

† 준 회원 : KAIST 전기 및 전자공학부 석사과정

†† 준 회원 : KAIST 전기 및 전자공학부 박사과정

††† 비 회원 : KAIST 전기 및 전자공학부 박사과정

†††† 종신회원 : KAIST 전기 및 전자공학부 정교수

Manuscript Received : July 7, 2016

First Revision : August 30, 2016

Accepted : August 30, 2016

\* Corresponding Author : Yong Man Ro(ymro@kaist.ac.kr)

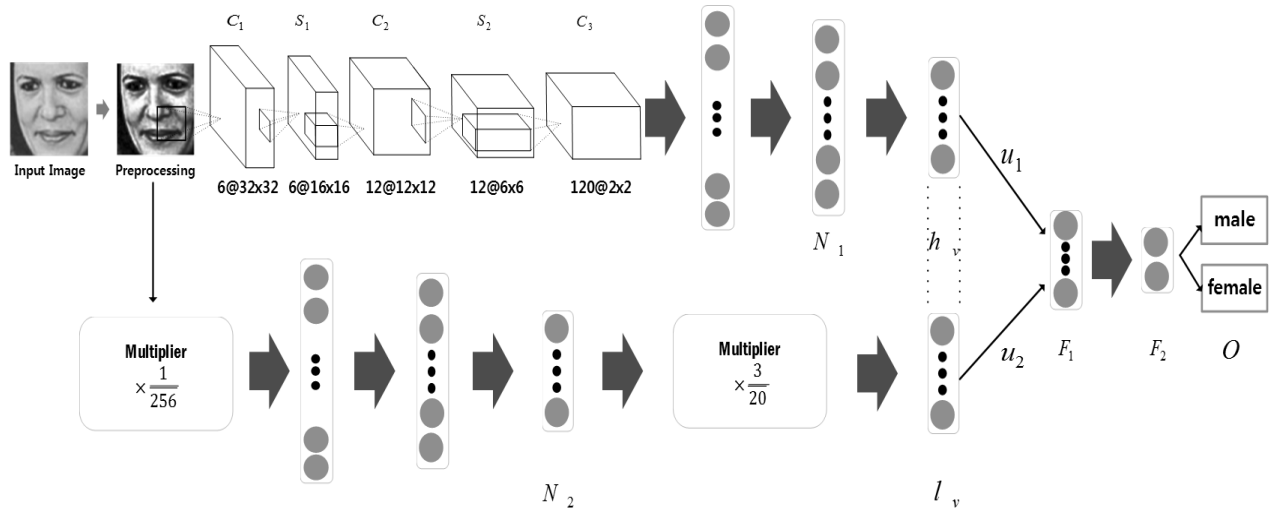


Fig. 1. Deep Learning Network for Gender Classification

지금까지 성별 인식을 향상시키기 위해 많은 연구가 이루어졌다. 가장 대표적인 방법으로 서포트 벡터 머신(SVM) 분류기를 이용한 성별 인식 방법[3]과, 아다부스트 (AdaBoost) 분류기를 이용한 방법[4] 등이 있으나 이러한 방법들은 영상 내 얼굴의 위치 및 조명 효과 및 얼굴 가림현상과 같은 외부 조건 그리고 얼굴 촬영 각도에 따라 인식이 크게 영향을 받기 때문에 아직 상업적으로 쓰이기에는 한계가 있다.

따라서 높은 성능을 가진 성별 인식기를 제작하기 위해선, 개개의 얼굴 영상 데이터로부터 성별 인식에 필요한 가장 지배적인 특징과 데이터에서 꼭 필요한 정보만을 찾아내는 방법이 요구된다. 최근에 ConvNets[5]와 같은 딥 러닝 모델이 이미지에서 상위 레벨 정보를 추출하는 문제에 대해 효율적인 방법임이 증명되었다. 마찬가지로 [6]은 딥 러닝 모델을 통해 이진영상 얼굴을 효과적으로 분류하는 방법을 제시하였고 [7]은 딥 러닝 네트워크를 깊게 만들어 각각의 얼굴 이미지에서 다양한 특징을 뽑아내 강인한 분류 성능을 제시하였다.

본 논문이 기여하는 점은 크게 두 가지이다.

먼저, LFW[8]와 같은 어려운 환경을 갖는 데이터 셋에서도 강인한 시스템을 개발하기 위해, 딥 러닝 기반의 실시간 성별 인식 방법을 제시한다. 제안하는 방법의 딥 러닝 구조는 Fig. 1에 제시된 바와 같이 성별인식에 필요한 얼굴의 골격과 같은 상위 단계 정보와 불필요한 정보가 제거된 하위 단계 정보를 적절한 비율로 연결된 구조로 이루어져 있다. 추가적인 성능향상을 위해 테스트 데이터로 성능 측정 시 각각의 이미지를 좌, 우로 평행 이동하여 3배로 증가시킨 후 더 많이 (2개 이상) 분류된 클래스를 원본 이미지의 최종 클래스로 결정하는 방법을 사용하였다.

두 번째, 딥 러닝 기반의 성별인식 방법을 임베디드 보드 환경에 구현하여 실시간 성별인식 시스템을 구축하였다. 실제 사물 인터넷의 경우, 낮은 전력 소비 및 기기의 부피를

최소화 하기 위해 PC 기반 제어가 아닌 임베디드 보드 기반의 맞춤형 제어가 이루어진다. 하지만 딥 러닝의 경우 수백만개 이상의 파라미터로 인해 대용량의 메모리가 요구된다. 그렇기 때문에 모바일 GPU 임베디드 보드 환경에서 정상적으로 동작할지 여부가 불투명하였다. 따라서 본 논문에서는 제안하는 성별인식 방법을 딥 러닝 라이브러리 중 하나인 Caffe[9]를 통해 구현한 후, 이를 임베디드 보드에 포팅하여, 사물 인터넷을 위한 실시간 성별인식 시스템을 구축하였다. 그 결과 소비전력을 PC 환경 대비 상당량 감소시켜 저전력 환경에서도 실제 성별 인식이 가능함을 입증하였다.

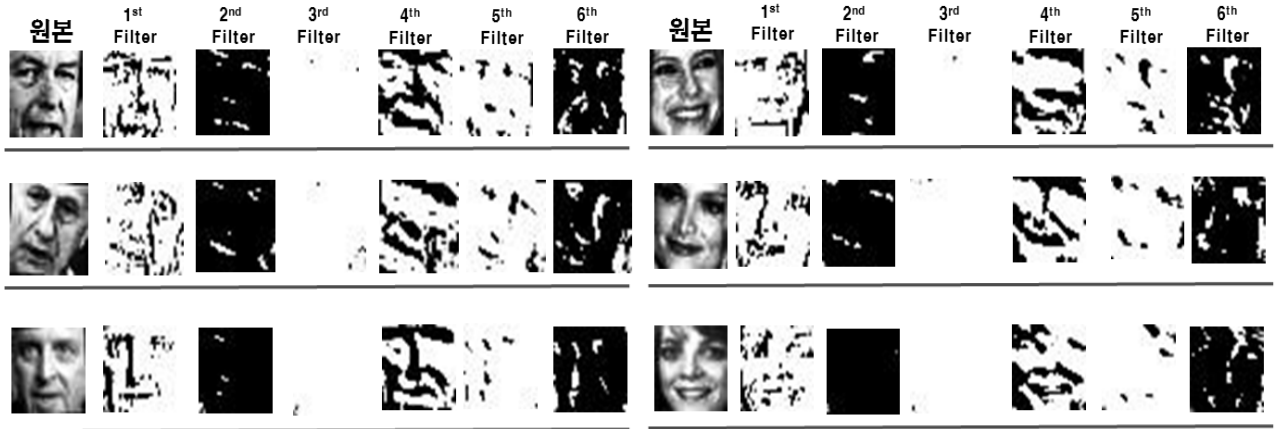
본 논문의 구성은 다음과 같다. 2장에서는 제안하는 성별인식 네트워크 및 알고리즘을 분석한다. 3장에서는 2장에서 제안한 알고리즘을 저전력 환경에서 구현하기 위해 임베디드 보드에 포팅하는 과정을 제시한다. 4장에서는 실험 환경 및 결과를 제시하고 마지막으로 5장에서는 결론을 제시한다.

## 2. 제안하는 성별인식 방법

본 논문에서는 이미지 데이터의 성별(예를 들어 여성:0, 남성:1)을 분류하는 것이 목표이다.

### 2.1 성별인식 시스템의 학습 전 이미지 전 처리

이미지 데이터를 학습하기 전, 몇 가지 전 처리 과정을 거친다. 먼저 [10]에서 제시한 얼굴 검출기를 통해 얼굴을 검출 한 후 이를 36x36사이즈로 이미지 크기를 변환한다. 그리고 사용한 data set에는 다양한 인종의 얼굴이 포함되어 있었고 다양한 촬영 환경(조명효과 등)이 반영되어 있기 때문에 이를 보정하기 위해 RGB 3 채널을 gray 1 채널로 변환 시킨 후 히스토그램 균등화 작업을 수행하였다.

Fig. 2. Feature Map Obtained from  $C_1$  Layer in Upper Network

2.2 딥 러닝 기반의 학습을 통해 상위 단계 정보와 하위 단계 정보를 최적화된 비로 연결한 딥 러닝 기반의 성별 인식 시스템

본 논문에서 제시하는 성별 인식 시스템은 Fig. 1에 제시되어 있고 크게 세 구조로 나뉘어진다[11]. 구체적인 내용은 다음과 같다.

첫 번째 구조는 얼굴의 윤곽과 같은 상위 단계 정보를 획득하기 위한 구조로서, 이미지와 같은 2차원 정보를 딥 러닝 학습 데이터로 사용할 수 있는 convolutional neural network (CNN)로 구성된다(본 논문에서는 이하 상위 네트워크라고 칭한다.) 상위 네트워크는 convolution layer와 활성화 함수 그리고 pooling layer로 구성되어 있으며 Fig. 1에서 상단에 위치한다. 먼저, convolution layer는 2차원 또는 3차원으로 구성된 데이터를 layer에서 학습을 통해 얻은 필터와 convolution 연산을 하여 얻어낸 특징을 추출하는 layer로, 입력된 이미지 데이터로부터 성별인식을 위해 의미있는 특징들의 2차원 또는 3차원 집합인 feature map을 뽑아내는 역할을 한다. 이후에는 출력 신호를 결정하는 활성화 함수를 사용한다. 본 논문에서는 이미지 데이터를 성별(0, 1) 즉, 2개의 라벨로 분류하기 때문에 각 feature map에서 1 이상의 매우 큰 수나, 0 이하의 매우 작은 수로 치우치는 것을 방지하고자 convolution layer 이후에는 출력 신호를 결정하는 활성화 함수로 sigmoid 함수 ( $\text{sigm}(\cdot)$ )를 사용하였다. 이는 Equation (1)과 같이 표현될 수 있다.

$$\sigma(x) = \text{sigm}(x) = \frac{1}{1 + e^{-x}}, \quad (1)$$

Equation (1)에서  $e$ 는 지수함수를 의미하고  $x$ 는 feature map에서의 단위 특징(픽셀) 값이다. 그리고 convolution layer를 거쳤을 때 각각의 단위 특징들이 갖는 값은 Equation (2)와 같이 표현된다.

$$a_j^l = \text{sigm}\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right), \quad (2)$$

Equation (2)에서  $w_{jk}^l$ 는 1-번째 layer의  $j$ -번째 feature map을 생성하는  $k$ -번째 가중치 행렬,  $a_j^l$ 는 1-번째 layer에서  $j$ -번째 feature map을 의미하고  $b_j^l$ 은 1-번째 layer에서  $j$ -번째 feature map에 걸리는 바이어스 항을 의미한다. 그리고 활성화 함수 이후에 있는 pooling layer의 경우 활성화 함수를 선정한 방법과 마찬가지로 최종 출력단 동작 영역을 0부터 1사이의 범위에 분포시키기 위해 패치 범위 안에 존재하는 값들을 평균화 하는 average pooling를 사용하였다. 이렇게 제안된 상위 네트워크를 통해 학습하면 Fig. 2와 같이 중간 layer에서 필터를 통해 다양한 얼굴 구조를 추출해 낼 수 있다. 딥 러닝 네트워크는 여기서 얻은 feature들을 기울기 하강방식(SGD) 기반의 학습방법인 back propagation을 진행하면서 비용 값을 수렴시키는 최적의 필터들을 찾아낸다.

하지만 상위 단계 정보만을 통해 성별을 인식할 경우 이미지에 포함되어 있는 노이즈나 다른 사물과 같은 불필요한 정보가 얼굴의 골격구조에 영향을 끼친다. 그렇기 때문에 첫 번째 구조를 보완하기 위해 성별 인식 시스템에 입력된 이미지의 불필요한 정보를 제거하는 기능을 갖는 두 번째 구조인 하위 네트워크가 사용된다.

하위 네트워크는 이미지 데이터에서 불필요한 픽셀 정보를 제거하는 기능을 갖고 Fig. 1에서 하단에 위치한 구조이다. 하위 네트워크는 딥 러닝의 가장 일반적인 구조인 DNN (deep neural network) 구조를 사용하였고 Equation (3)과 같이 표현된다.

$$f_{\text{output}}(x_{\text{input}}) = w^T x_{\text{input}} + b, \quad (3)$$

여기서,  $w$ 는 가중치 벡터를 의미하고,  $x_{\text{input}}$ 는 입력 벡터를,  $b$ 는 바이어스 벡터를 의미하며,  $T$ 는 벡터 전치(transpose)를

의미한다. 그리고  $f_{output}$ 은 DNN의 출력 벡터를 의미한다.

입력 이미지(36x36)를 벡터화시켜 1296 차원의 벡터로 만든 후 2번의 이전 layer의 특징들과 현재 layer의 특징들이 상호 연결되어 있는 fully connected layer를 2번 거친 후 벡터의 차원을 축소하여(1296차원 → 240차원 → 120차원) 최종적으로 학습에 필요한 120개의 정보를 뽑아낸다. 한 개의 fully connected layer만 생성하여 1296개의 정보를 120개의 정보로 한 번에 줄일 경우 급격한 정보 손실로 인해 학습이 원활이 이루어지지 않아 2개의 fully connected layer를 삽입하였다. 또한 하위 네트워크에서 fully connected layer만을 사용하여 학습할 경우, 하위 네트워크 결과 값이 상위 네트워크 결과 값 보다 매우 큰 값이 나온다. 하위 네트워크는 상위 네트워크를 보조해주는 역할로 사용되어야 하므로 하위 네트워크 앞단에 값의 크기를 줄여주는 곱셈기를 삽입하였고 학습된 비용값이 전역 최소값에 수렴하도록 실험적 방법을 통해 하위 네트워크 끝단에 추가적인 곱셈기를 삽입하였다.

마지막으로 세 번째 구조는 상위 단계 정보를 파악하는 상위 네트워크로부터 얻어진 벡터와 세부 정보를 파악하는 하위 네트워크로부터 얻어진 벡터를 하나의 특징 벡터로 만드는 연결 네트워크이며 Fig. 1에서 오른쪽 부분이다. 두 구조를 결합하기 위해 상위 네트워크의 출력단을 벡터화 시킨 후 이를 하위 네트워크의 출력단에 생성된 벡터와 연결하여 Equation (4)와 같이 표현할 수 있다.

$$F_1 = \text{sigm}(u_1^\top h_v + \beta u_2^\top l_v + b), \quad (4)$$

Equation (4)에서  $u_1$ 은 상위 네트워크의 가중치를,  $u_2$ 는 하위 네트워크의 가중치를 나타내고  $h_v$ 와  $l_v$ 는 각각 상위 네트워크와 하위 네트워크를 표현하는 벡터이다. 그리고  $\beta$ 는 상위 네트워크에 대한 하위 네트워크의 비율을 조절하기 위한 가중치 상수로,  $0 \leq \beta \leq 1$ 의 범위를 갖는다. 연결 네트워크도 하위 네트워크와 마찬가지로 정보의 급격한 손실을 방지하기 위해 출력단 앞에 2개의 fully connected layer를 삽입하였다. 이때  $F_1$  layer와  $F_2$  layer의 출력과 연산되는 가중치 벡터를 각각  $k_{F_1}$ ,  $k_{F_2}$ 라 하고 최종 출력단 벡터를  $O$ 라고 정의하면, 최종 출력  $O$ 는 Equation (5)와 같이 표현할 수 있다.

$$O = k_{F_2}^\top (k_{F_1}^\top F_1), \quad (5)$$

### 2.3 크로스 엔트로피 비용함수를 이용한 성별 인식 시스템의 학습과정

본 논문에서는 이미지 데이터를 학습하기 위한 비용함수로 Equation (6), (7)과 같이 크로스 엔트로피 함수를 사용

한다. 비용함수를 결정하는 매개변수로는 가중치 행렬( $W$ ), 바이어스( $b$ ),  $n$ 번째 학습(epoch)에서 상위 네트워크와 하위 네트워크의 비율 조절 상수( $\beta_n$ )로 구성된다. 이 중에서 상위 네트워크와 하위 네트워크는 각각 얼굴 전반적인 특징과 얼굴의 세부적인 특징을 지칭하기 때문에 두 네트워크의 가중치를 조절해주는  $\beta_n$ 이 비용함수에 가장 지배적인 영향을 끼친다. 그렇기 때문에 Equation (6), (7), (8)을 이용하여 비용함수를 최소화시키는  $\beta$ 값을 찾는 것이 학습 과정의 핵심이다.

$$p_j = \frac{e^{O_j}}{e^{O_0} + e^{O_1}}, \quad j=0 \text{ or } 1, \quad (6)$$

Equation (6)에서  $p_j$ 는 각 성별의 소프트맥스 확률이다. 그리고  $j = 0$  or  $1$ 이고,  $0$ 은 여성일 때,  $1$ 은 남성일 때를 의미한다. 또한  $O_0$ 은 여성일 때의 출력 값,  $O_1$ 은 남성일 때의 출력 값이다.

Equation (6)에서 구한 소프트맥스 확률을 이용하여 Equation (7)과 같이 비용함수를 표현할 수 있다. 여기서  $L$ 은 비용값이며,  $o_j$ 는 라벨이다.

$$L = - \sum_{j=gender} o_j \log p_j \quad (7)$$

Equation (8)에서 가장 최적화된 네트워크를 생성하는 비율조정 상수  $\beta$ 를 찾기 위해 확률 기울기 하강 방식을 이용한다.  $\eta$ 를 반복적인 학습에서 다음 학습 기준점을 어느 정도로 옮길지 척도가 되는 값인 learning rate라고 할 때 ( $n+1$ )번째 학습에서의 비율 조절 상수는 Equation (9)와 같이 표현할 수 있다. 비용함수의 최적의 해를 구할 때까지 Equation (9)를 반복해준다.

$$\underset{\beta}{\text{argmin}} L(W, b, \beta) \rightarrow \underset{\beta}{\text{arg}} \frac{\partial L(W, b, \beta)}{\partial \beta} = 0, \quad (8)$$

$$\beta_{n+1} = \beta_n - \eta \frac{\partial L(W, b, \beta)}{\partial \beta}, \quad 0 \leq \beta_n, \beta_{n+1} \leq 1, \quad (9)$$

### 2.4 데이터 증가 방식을 이용한 성별 인식 시스템의 테스트 과정

본 논문에서는 이전에 제안한 방법[11]과 달리 테스트 과정의 경우 학습과정에 사용된 네트워크에 데이터를 증가시키는 알고리즘을 Fig. 3과 같이 추가로 삽입하였다. 얼굴이 crop된 데이터의 경우 이미지의 배경, 특이한 얼굴모양과 같은 다양한 요인으로 인해 얼굴이 완벽하게 정렬되지 않을 수 있다. 그렇기 때문에 해당 데이터들을 좌, 우로 조금씩

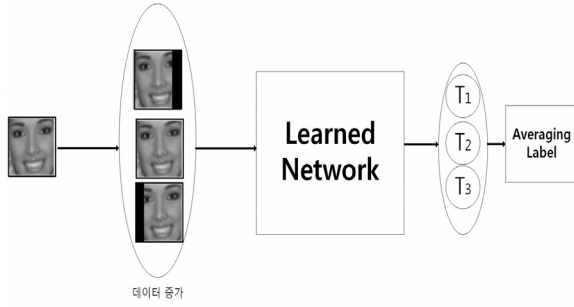


Fig. 3. Network for Test Dataset

평행 이동하거나 좌, 우로 약간씩 회전하여 나온 출력 값들을 이용하여 라벨을 구한 후 더 빈번하게 발생한 라벨의 성별을 해당 테스트용 데이터의 성별로 지정한다면 정렬문제로 인한 문제점을 보정할 수 있다.

그래서 본 논문에 적용되는 성별인식 알고리즘은 출력된 각각의 라벨(0 또는 1)들의 평균값을 취하여 0.5 이상이면 여성으로, 0.5 이하면 남성으로 분류하는 방법을 추가하였다. 원본 테스트 이미지의 라벨을 T1이라고 하고 왼쪽, 오른쪽으로 평행이동 시킨 이미지의 출력 라벨을 T2, T3라고 하면 최종 성별 G는 Equation (10)과 같이 표현할 수 있다.

$$G = \begin{cases} 0, & \text{if } \frac{T_1 + T_2 + T_3}{3} < 0.5 \\ 1, & \text{if } \frac{T_1 + T_2 + T_3}{3} \geq 0.5 \end{cases}, \quad (10)$$

성별 인식 네트워크가 여성으로 분류하면 출력 값으로 0을, 남성으로 분류하면 출력 값으로 1을 갖는다. 실제로 실험 과정에서 데이터를 증가시키기 위해 얼굴을 평행이동만 해서 데이터 수를 3배로 증가시킨 경우, 회전만 해서 데이터 수를 3배로 증가시킨 경우 그리고 평행이동과 회전을 모두 해서 데이터 수를 5배로 증가시킨 경우를 모두 실험해 보았다. 그 결과 Table 1과 같이 데이터를 평행이동만 하여 데이터 수를 증가 시킨 경우에서 가장 높은 인식률을 보였다.

Table 1. Gender Classification Accuracy According to Data Augmentation Method

	Parallel Translation	Rotation	Parallel Translation + Rotation
Accuracy	91.3%	90.7%	91.1%

### 3. 모바일 GPU 환경에서의 딥 러닝 기반 성별 인식 시스템 구현

임베디드 보드는 특정 기능만 수행하도록 제작된 보드이므로 PC에 비해 메모리 용량이나 CPU성능이 떨어진다. 그렇기 때문에 일반적으로 GPU에 프로그램을 이식할 때는 PC에서 프로그램을 구현한 후 PC와 임베디드 보드를 연결해서 프로그램을 포팅하거나 이동식 저장장치를 통해 프로그램을 옮긴 후 포팅한다.

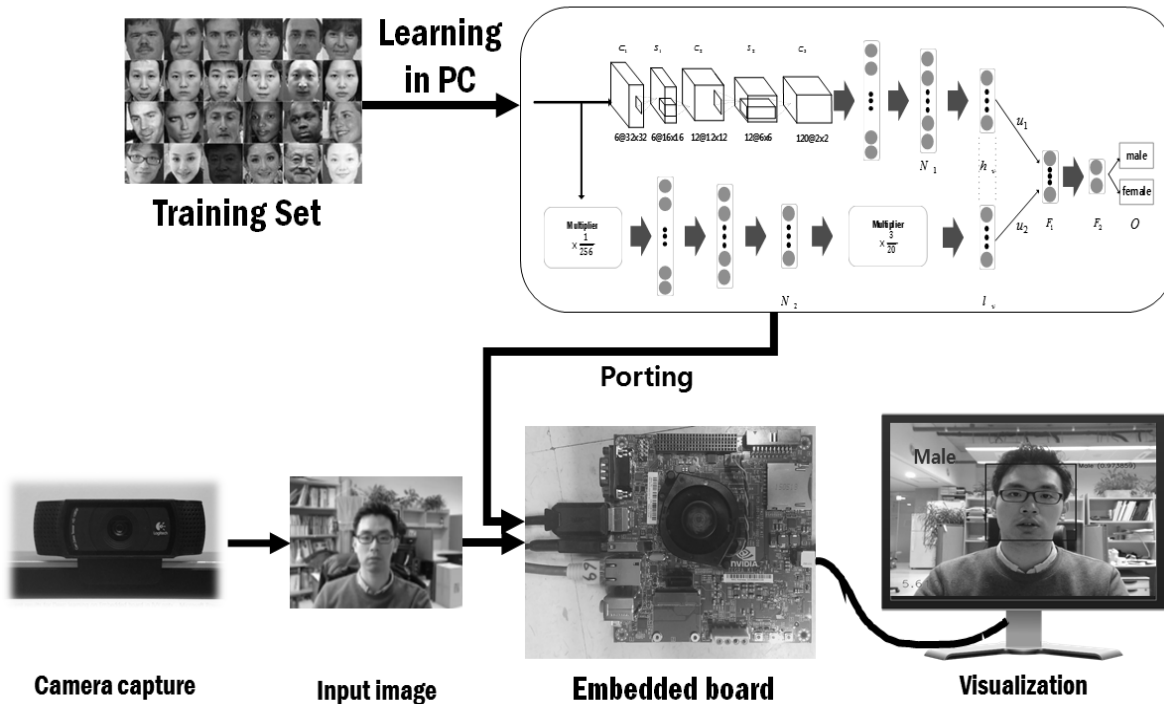


Fig. 4. Flow chart of deep learning based gender classification system on embedded board

Fig. 4는 임베디드 보드환경에서 본 논문에서 제안하는 딥 러닝을 이용한 성별인식 시스템의 개요도이다. 먼저, 인식할 시스템을 PC환경 C++에서 학습시킨다. 그리고 임베디드 보드의 경우 상당수가 리눅스 기반 운영체제에서 동작하기 때문에 make file도 함께 제작해준다. 그리고 보드 세팅을 위해 보드를 랜(LAN) 선과 연결 후 NVIDIA installer를 설치 후 GPU를 작동시키기 위한 Cuda SDK를 설치한다. 그리고 딥 러닝을 위해 필요한 라이브러리인 Caffe와 영상 처리에 필요한 opencv 라이브러리를 패키지를 설치한다. 이후 PC에서 구현한 성별인식 시스템과 make file을 보드에 포팅하고 마지막으로 임베디드 보드와 카메라 캡처 시스템을 직접 연결하여 성별인식 시스템을 구현하였다.

#### 4. 실험 결과 및 분석

본 논문에서 제안하는 성별인식 시스템의 인식률을 검증하기 위해 matlab에서 공용 데이터를 이용해 성별인식률을 측정하였다. 그리고 이를 실시간으로 동작시켜 PC와 임베디드 보드 환경에서 각각 C++로 구현된 프로그램으로 전력 측면에서 얼마나 이득이 있는지 측정하였다.

##### 4.1 사용된 데이터 및 성별 인식률

본 논문에서는 제안하는 네트워크를 구현하기 위해 Caffe 라이브러리를 사용하였다. 그리고 제안하는 알고리즘의 성능을 가하기 위해 LFW[8]와 CACD[12] 를 이용하였다. 학습용 데이터를 구성하기 위해 혼합된 dataset에서 남, 여 각각 7,500장을 랜덤 추출하였다. 그리고 dataset 숫자를 증가시키기 위해 각각의 이미지를 좌우반전과 [-2, 0, 2] 픽셀만큼

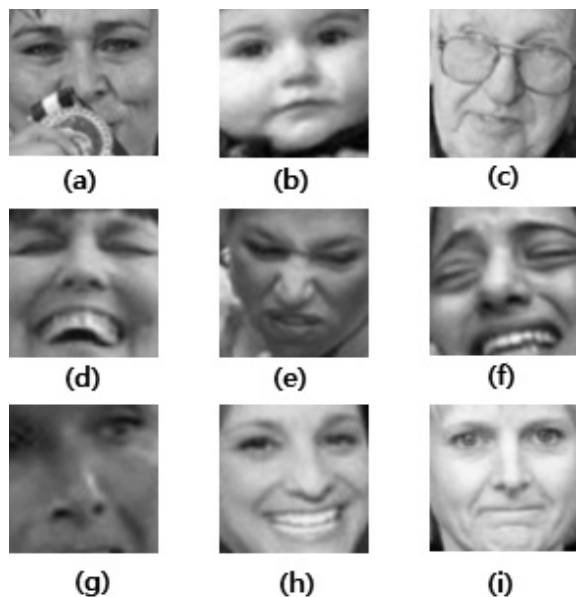


Fig. 5. Some Examples of Mis-classification

평행이동 시켜 학습용 데이터 량을 총 6배로 증가시켰다. 테스트용 데이터로는 학습과정에 사용되지 않은 것들 중 남, 여 각각 500장을 임의로 추출하여 사용하였다.

테스트용 데이터를 좌우 2픽셀씩 평행이동 해서 라벨을 얻은 경우, 좌, 우 3도씩 회전해서 라벨을 얻은 경우, 평행이동과 회전 모두 해서 라벨을 얻은 경우 각각에 대한 인식률은 Table 1과 같다. 테스트 과정에서 데이터를 평행이동만 하여 수를 증가시키는 경우 가장 높은 인식률을 보였다.

Fig. 5는 오분류된 몇 가지 사례들을 제시하고 있다. (a)는 다른 물체에 의해 얼굴이 가려진 경우, (b), (c)는 피 촬영자가 너무 어리거나 고령인 경우이다. (d), (e)는 독특한 표정을 짓고 있는 경우, (f)와 (g)는 얼굴 crop이 비정상적으로 진행된 경우이다. (h)와 (i)는 각각 강한 미소로 인한 치아와 창백한 입술 색으로 성별을 오분류된 사례이다.

##### 4.2 임베디드 환경 및 PC 실험과의 비교 분석

제안하는 알고리즘을 포팅 한 임베디드 보드는 NVIDIA에서 제작한 Jetson TK1이고 Ubuntu 운영체제 기반으로 동작 한다. 그리고 CPU, GPU, 메모리를 하나의 칩으로 패키징 한 Tegra SoC가 Jetson TK1에 장착되어 있기 때문에 별도의 PC 없이 학습된 네트워크를 보드에 포팅이 가능하다. 그리고 실험에 사용된 PC는 Intel(R) Core i7-4790k CPU를 사용하고 메모리 용량은 32GB이다. 그리고 GPU로 GeForce GTX TITAN X를 사용하였다.

PC에서 학습된 네트워크를 임베디드 보드에 가져왔기 때문에 인식 정확도는 그대로여야 하지만 PC 그래픽카드와 휴대용 임베디드 보드의 그래픽 카드 성능 차이로 인해 실제 시스템을 구현 시 소비전력과 초당 처리 가능한 이미지 수가 Table 2와 같이 차이가 존재한다. 본 논문의 저자가 사용하는 PC를 기준으로 비교 했을 때 임베디드 보드에서 처리 가능한 이미지 수는 약 2~3배 정도 적지만 소비전력 측면에서 50배 이상 에너지가 절감되는 이점이 있었다.

Table 2. Power Consumption and Processing Speed When Real-Time Gender Classification System is Operated in PC and Embedded PC

	PC	Embedded Board
Power Consumption (Watts)	560	10.5 [14]
Processing Speed (fps)	12.25	4.9

#### 5. 결론

본 논문에서는 성별 인식을 위해 딥 러닝을 이용한 새로운 알고리즘을 제안하였다. CNN 구조를 이용해 필터를 통한 다양한 얼굴 골격을 추출하고 DNN을 이용해 세부적인

정보를 추출한다. 그리고 두 종류의 정보를 적절한 비로 결합하면서 feature 값을 0과 1 사이로 유지하면서 정보 손실량을 최소화 하였다. 그리고 test과정에서 얼굴 alignment 문제를 보정하기 위해 하나의 이미지를 평행이동을 통해 세장으로 증가시켜 성능을 측정하였다. 그 결과 91.3%의 성능을 보였다.

그리고 위에서 제시한 알고리즘을 C++를 이용하여 실시간 프로그램으로 구현한 후 Jetson TK1 보드에 임베디드 포팅을 하고 PC환경과 비교를 하였다. 그 결과 임베디드 보드에서 초당 처리할 수 있는 이미지 수는 PC에 비해 적었지만 소비전력 측면에서 수십 배의 이점이 있어 향후 사물인터넷이 적용된 제품에 사용자의 정보를 얻기 위해 다양하게 사용될 수 있을 것으로 보인다.

### References

[1] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015.

[2] C. Perez, J. Tapia, P. Estévez, and C. Held, "Gender classification from face images using mutual information and feature fusion," *International Journal of Optomechatronics*, Vol.6, No.1, pp.92-119, 2012.

[3] B. Moghaddam and M. H. Yang, "Learning gender with support faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, No.5, pp.707-711, 2002.

[4] S. Baluja and H. A. Rowley, "Boosting sex identification performance," *International Journal of Computer Vision*, Vol.71, No.1, pp.111-119, 2007.

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, Vol.86, No.11, pp.2278-2324, 1998.

[6] Y. Sun, X. Wang, and X. Tang, "Hybrid deep learning for face verification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013.

[7] D. H. Kim, S. T. Kim, and Y. M. Ro, "Latent feature representation with 3-D multi-view deep convolutional neural network for bilateral analysis in digital breast tomosynthesis," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2016.

[8] G. B. Huang, M. Ramesh, T. Berg, and E. Learned\_miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Technical report 07-49, UMass, 2007.

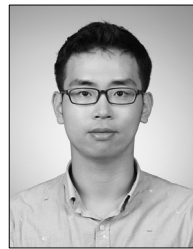
[9] Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding [Internet], <http://caffe.berkeleyvision.org/>, 2013.

[10] Gary Bradsky and Adrian Kaebler, "Viola Jones Face Detector," in *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 1st edition.

[11] Hyunwook Jeong, Dae Hoe Kim, Wisam J. Baddar, and Yong Man Ro, "Real-time Gender Classification based on Deep Learning in Embedded System," in *the Proceedings of the KIPS Symposium*, 2016.

[12] Wen, Yandong, Zhifeng Li, and Yu Qiao, "Latent Factor Guided Convolutional Neural Networks for Age-Invariant Face Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[13] NVIDIA developer blog [Internet], <https://devblogs.nvidia.com/parallelforall/embedded-machine-learning-cudnn-deep-neural-network-library-jetsot-tk1/>.



정 현 욱

e-mail : sceinstein@kaist.ac.kr  
 2015년 고려대학교 전기전자전파공학부 (학사)  
 2015년~현 재 KAIST 전기 및 전자공학부 석사과정  
 관심분야 : Deep Learning & Computer Generated Holography



김 대 회

e-mail : dhkim10@kaist.ac.kr  
 2010년 한양대학교 전자통신컴퓨터공학부 (학사)  
 2012년 KAIST 전기 및 전자공학부(석사)  
 2012년~현 재 KAIST 전기 및 전자공학부 박사과정  
 관심분야 : Deep Learning & Medical Imaging(Mammography, 3D Ultrasound, Digital Breast Tomosynthesis)



Wisam J. Baddar

e-mail : wisam.baddar@kaist.ac.kr  
 2010년 University of Jordan, School of Electrical Engineering(학사)  
 2014년 KAIST 전기 및 전자공학부(석사)  
 2014년~현 재 KAIST 전기 및 전자공학부 박사과정  
 관심분야 : Machine Learning, Deep Learning, Facial Analysis & Medical Imaging



노 응 만

e-mail : ymro@kaist.ac.kr

1985년 연세대학교 전자공학과(학사)

1987년 KAIST 전기 및 전자공학과(석사)

1992년 KAIST 전기 및 전자공학과(박사)

1987년~1988년 Columbia University

연구원

1996년~1997년 University of California, Berkely 연구원

2006년~2007년 University of Toronto 방문교수

1997년~현 재 KAIST 전기 및 전자공학과 정교수

관심분야: Deep Learning and Machine Learning on Image

Processing and Computer Vision, High-performance

Object Recognition, Emotion Recognition, Medical

Image Processing, 3D Processing, Quality Measureme