일반논문 (Regular Paper)
방송공학회논문지 제24권 제6호, 2019년 11월 (JBE Vol. 24, No. 6, November 2019)
https://doi.org/10.5909/JBE.2019.24.6.1113
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

합성곱 신경망 기반 밝기-색상 정보를 이용한 얼굴 위변조 검출 방법

김 은 석[a)], 김 원 준[a)‡]

Face Anti-Spoofing Based on Combination of Luminance and Chrominance with Convolutional Neural Networks

Eunseok Kim[a)] and Wonjun Kim[a)‡]

요 약

본 논문에서는 얼굴의 밝기와 색상 정보를 함께 이용한 합성곱 신경망 기반의 얼굴 위변조 검출 방법을 제안한다. 제안하는 방법은 적층된 합성곱 신경망과 보조 신경망을 이용하여 실제 얼굴과 위변조된 얼굴의 밝기 특징과 색상 특징을 독립적으로 추출한다. 기존의 방법과는 달리, 본 논문에서는 추출된 특징을 단순 결합(Concatenation)하는 것이 아니라 주의 모듈(Attention Module)을 이용하여 적응적(Adaptively)으로 조합할 수 있도록 하였다. 또한, 효과적인 분류기 학습을 위하여 대비 손실함수(Contrast Loss Function)를 새롭게 제안하였는데, 대비 손실함수는 동일 클래스 내의 특징 간의 차이는 최소화 시키고 서로 다른 클래스의 특징 간의 차이는 최대화 시킴으로써 특징의 분별력을 높인다. 다양한 실험을 통해 본 논문에서 제안하는 방법이 기존 얼굴 위변조 검출 방법 대비 개선된 성능을 보임을 확인하고 그 결과를 분석한다.

Abstract

In this paper, we propose the face anti-spoofing method based on combination of luminance and chrominance with convolutional neural networks. The proposed method extracts luminance and chrominance features independently from live and fake faces by using stacked convolutional neural networks and auxiliary networks. Unlike previous methods, an attention module has been adopted to adaptively combine extracted features instead of simply concatenating them. In addition, we propose a new loss function, called the contrast loss, to learn the classifier more efficiently. Specifically, the contrast loss improves the discriminative power of the features by maximizing the distance of the inter-class features while minimizing that of the intra-class features. Experimental results demonstrate that our method achieves the significant improvement for face anti-spoofing compared to existing methods.

Keyword : face anti-spoofing, luminance and chrominance, convolutional neural networks, attention module, contrast loss

a) 건국대학교 전기전자공학부(Department of Electrical and Electronics Engineering, Konkuk University)
‡ Corresponding Author : 김원준(Wonjun Kim)
E-mail: wonjkim@konkuk.ac.kr
Tel: +82-2-450-3396
ORCID: https://orcid.org/0000-0001-5121-5931

※ 이 논문은 2017년도 정부(과학기술정통부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2017R1C1B2003044).
※ This paper was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2017R1C1B2003044).
· Manuscript received August 29, 2019; Revised September 18, 2019; Accepted September 18, 2019.

Ⅰ. 서 론

최근 스마트폰, 태블릿 PC와 같은 다양한 모바일 기기가 널리 보급되면서, 고성능 보안 시스템에 대한 수요 또한 증가하고 있다. 사용자가 직접 인증을 위한 키(Key)를 휴대하거나, 비밀번호 혹은 패턴 등을 이용하는 전통적인 사용자 인증 방법은 분실이나 망각의 우려가 있어 기업과 개인의 보안 수요를 충족시키지 못하고 있다. 이와 달리 생체 기반의 인증 시스템은 개개인의 고유한 신체적 특징을 이용하여 사용자를 식별하기 때문에 분실 및 망각의 위험이 없고, 사용이 편리하다는 장점이 있어 이와 관련된 기술 연구가 활발히 진행되고 있다. 지문, 얼굴, 홍채, 정맥 등 생체 인증 시스템에서 사용되는 다양한 생체 특징 중에서도 얼굴은 카메라만을 이용하여 접촉 없이 쉽게 영상을 획득할 수 있다는 장점이 있어 최근 보안 영역에서 가장 널리 사용되고 있다[1]. 그러나 얼굴 인증 시스템은 사용자의 얼굴이나 동영상을 기반으로 한 위변조 공격에 취약한 단점이 있다. 이러한 문제점을 극복하기 위하여 얼굴 움직임 정보, 영상 화질 정보, 얼굴의 질감(Texture) 정보 등을 이용하여 얼굴 위변조 공격을 검출하려는 시도가 다양한 연구자들에 의해 이루어지고 있다.

초기에는 눈의 깜빡거림, 입의 움직임, 머리의 회전 등 움직임 특성을 이용하여 얼굴 위변조 여부를 판별하는 방법이 연구되었다[2][3][4]. 이와 같은 방법은 사진 위변조 공격 등 움직임이 없는 위변조 공격에는 효과적으로 대응할 수 있으나, 동영상 위변조 공격과 같이 움직임 정보를 포함하고 있는 위변조 공격에는 취약하다는 단점이 있다. 최근에는 실제 얼굴 영상과 위변조된 얼굴 영상의 화질 차이를 기반으로 하는 얼굴 위변조 검출 방법도 연구되었다. 이러한 방법은 주로 카메라를 통해 획득된 영상의 화질 평가 결과를 이용하여 위변조 검출을 수행한다[5]. 이와 같은 방법은 생체 정보의 종류에 관계없이 높은 검출 정확도를 보이지만, 고화질의 위변조 얼굴 영상을 이용한 공격에는 취약하다는 한계가 있다. 한편으로, 얼굴의 질감 특성 차이를 이용하는 방법도 연구되었다. 이 방식은 주로 국부 이진 패턴(Local Binary Pattern, LBP)[6]과 같은 국부 영상 표현자(Local Image Descriptor)를 이용하여 실제 얼굴 영상과 위변조된 얼굴 영상의 질감 특성 차이를 표현한다[7][8]. 질감 기반의 방법은 구현이 쉽고 검출 시간이 짧다는 장점으로 인하여 관련 연구가 활발히 진행되어 왔으나, 다양한 환경에서 획득되는 잡음이나 위변조 과정에서 발생하는 불균일한 영상 변화에 취약하다는 단점이 있다. 가장 최근에는 심층 신경망(Deep Neural Network, DNN), 특히 합성곱 신경망(Convolutional Neural Network, CNN)의 성공에 힘입어 이를 얼굴 위변조 검출에 적용하고자 하는 시도가 활발히 진행되고 있다[9][10]. 이러한 방법은 학습을 통하여 분별력 있는 특징을 효과적으로 도출할 수 있어, 기존의 질감 기반의 위변조 검출 방법 대비 향상된 성능을 보여주었다.

이처럼 다양한 연구를 통해 얼굴 위변조 검출 분야에서 많은 발전이 이루어졌지만, 기존의 방법은 주로 얼굴 영상의 밝기 정보에만 집중하며, 얼굴 위변조 여부 판별에 밝기 정보와 마찬가지로 유용할 수 있는 색상 정보는 간과하고 있다는 한계가 있다. 따라서 얼굴 영상의 색상 정보도 함께 고려하기 위해 국부 이진 패턴을 이용하여 얼굴 영상의 밝기 공간과 색상 공간으로부터 질감 특징을 독립적으로 추출하는 방법이 제안되었다[14]. 그러나 이러한 방법은 여전히 사용된 영상 표현자에 의해 성능이 좌우되는 한계가 있다.

본 논문에서는 이러한 한계점을 극복하기 위하여 얼굴의 밝기와 색상 정보를 함께 이용한 합성곱 신경망 기반의 얼굴 위변조 검출 방법을 제안한다. 제안하는 방법은 적층된 합성곱 신경망 구조를 이용하여 실제 얼굴과 위변조된 얼굴로부터 밝기 특징과 색상 특징을 각각 추출하는데, 대량의 데이터로부터 학습을 통해 특징을 추출하므로 다양한 위변조 공격 시나리오에 효과적으로 대응할 수 있다. 기존의 방법과는 달리, 본 논문에서는 추출된 밝기 특징과 색상 특징을 단순 결합하는 것이 아니라 주의 모듈(Attention Module)을 통해 특징 간의 관계를 고려함으로써 적응적으로 특징을 조합할 수 있도록 하였다. 또한, 동일 클래스 내 특징 간의 차이는 최소화 시키고, 동시에 서로 다른 클래스 내 특징 간의 차이는 최대화 시켜 특징의 분별력을 높일 수 있는 대비 손실함수를 제안함으로써 더욱 효과적인 분류기 학습이 가능하도록 하였다. 먼저 2장에서는 제안하는 합성곱 신경망 구조, 사용된 주의 모듈 및 손실함수에 대해 자세히 설명하고, 3장에서는 실험을 통해 제안하는 방법이 기존의 방법 대비 성능이 개선됨을 보인다. 마지막으로 4장

에서 결론으로 본 논문을 마무리한다.

Ⅱ. 제안하는 방법

제안하는 방법은 크게 특징을 추출하는 단계와 추출된 특징을 재조합하는 단계, 마지막으로 추출된 특징을 이용하여 얼굴 위변조 검출을 수행하는 단계로 나누어진다. 영상을 표현하는 다양한 색상 모델 중에서도 YCbCr 색상 모델은 영상의 밝기 정보(Y)와 색차 정보(Cb, Cr)를 분리하여 표현할 수 있으므로, 본 논문에서는 영상의 밝기 공간과 색상 공간으로부터 독립적으로 특징을 추출하기 위하여 YCbCr 색상 모델을 사용하였다. 제안하는 방법은 합성곱 신경망 구조를 이용하여 얼굴 영상의 Y, Cb, Cr 공간으로부터 각각 특징을 추출하고, 추출된 특징들을 주의 모듈을 이용하여 적응적으로 재조합한다. 최종적으로 손실함수와 재조합된 특징을 이용하여 분류기를 학습하고 생성된 모델을 바탕으로 얼굴 위변조 검출을 수행한다. 제안하는 전체 합성곱 신경망 구조는 그림 1과 같다.

1. 합성곱 신경망을 이용한 밝기 특징과 색상 특징 추출

위에서 언급한 것과 같이, 가장 먼저 얼굴 영상으로부터 밝기 특징과 색상 특징을 독립적으로 추출하기 위하여 합성곱 신경망 구조가 사용되었다. 제안하는 방법은 적층된 합성곱 신경망 구조를 이용하여 입력 얼굴 영상의 Y, Cb, Cr 공간으로부터 특징을 추출하여 밝기 정보와 색상 정보를 함께 고려함으로써 효과적인 얼굴 위변조 판별이 가능하도록 하였다. 특징 추출을 위하여 사용된 합성곱 신경망 구조는 표 1과 같이 세 개의 세부 합성곱 신경망으로 구성되어 있으며, 각 세부 합성곱 신경망의 구조는 모두 동일하다. 각 세부 합성곱 신경망은 다섯 개의 합성곱 계층(Convolution Layer)과 다섯 개의 최대 풀링 계층(Max-Pooling

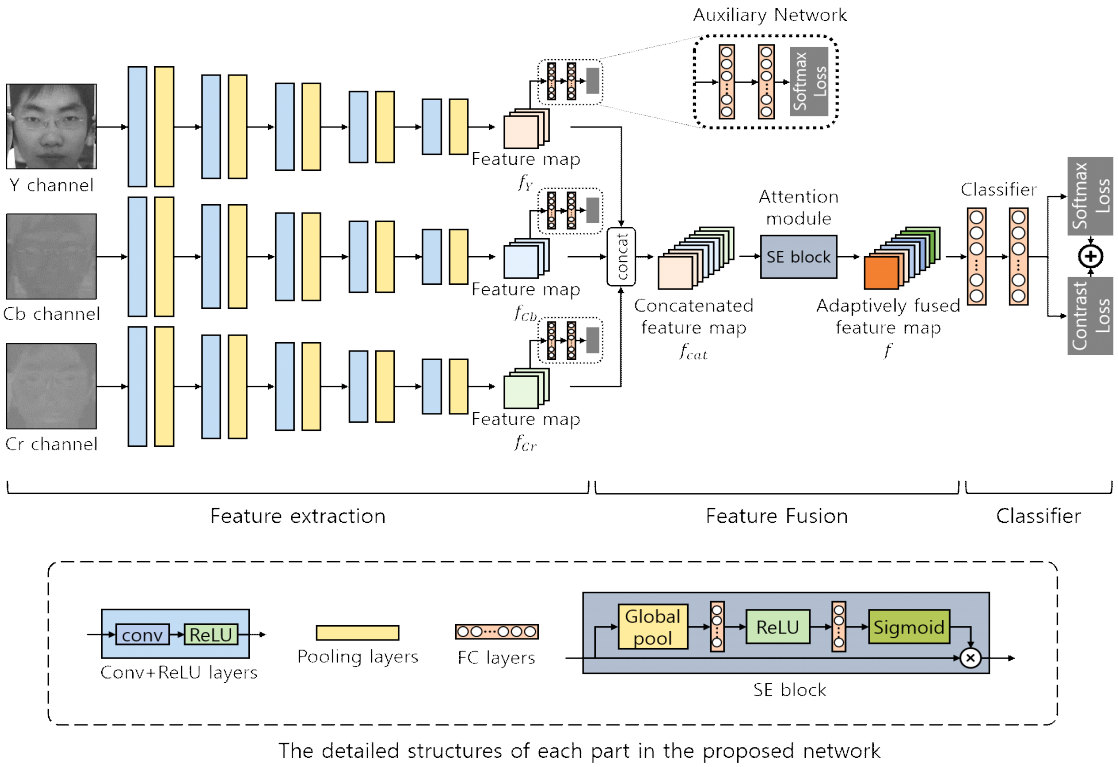

그림 1. 제안하는 합성곱 신경망 구조 개요
Fig. 1. Overview of the proposed convolutional neural network architecture

표 1. 본 논문에서 특징 추출을 위해 사용한 세부 합성곱 신경망 구조 CNN_Y, CNN_{Cb}, CNN_{Cr}. conv(w, s, N) 은 스트라이드(stride) s와 w×w 픽셀 크기를 가진 N개의 필터가 있는 합성곱 계층을 나타낸다. pool(w, s) 은 스트라이드 s, w×w 픽셀 크기의 최대 풀링 계층을 의미한다.

Table 1. Details of the network architecture used in this paper. conv(w, s, N) denotes convolution layer which has N filters of size w×w with stride s. pool(w, s) is a w×w max-pooling layer with stride s;ReLU is used as the non-linear activation function.

	CNN for Y space (CNN_Y)	CNN for Cb space (CNN_{Cb})	CNN for Cr space (CNN_{Cr})
Input	112×112×1	112×112×1	112×112×1
Conv-1	conv(3, 1, 32)	conv(3, 1, 32)	conv(3, 1, 32)
Pooling-1	pool(3, 2)	pool(3, 2)	pool(3, 2)
Conv-2	conv(3, 1, 64)	conv(3, 1, 64)	conv(3, 1, 64)
Pooling-2	pool(3, 2)	pool(3, 2)	pool(3, 2)
Conv-3	conv(3, 1, 128)	conv(3, 1, 128)	conv(3, 1, 128)
Pooling-3	pool(3, 2)	pool(3, 2)	pool(3, 2)
Conv-4	conv(3, 1, 256)	conv(3, 1, 256)	conv(3, 1, 256)
Pooling-4	pool(3, 2)	pool(3, 2)	pool(3, 2)
Conv-5	conv(3, 1, 512)	conv(3, 1, 512)	conv(3, 1, 512)
Pooling-5	pool(3, 2)	pool(3, 2)	pool(3, 2)

Layer)으로 구성되어 있고, ReLU(Rectified Linear Unit) 함수가 모든 합성곱 계층 뒤에 추가되었다.

먼저, 얼굴 영역 영상 I를 Y, Cb, Cr 영상, 즉 I_Y, I_{Cb}, I_{Cr}으로 변환한다. 그리고, 세 개의 세부 합성곱 신경망 CNN_Y, CNN_{Cb}, CNN_{Cr}을 이용하여 각 공간으로부터 특징맵(Feature Map) f_Y, f_{Cb}, f_{Cr}을 추출한다. 이러한 특징 추출 과정은 다음과 같이 표현할 수 있다.

$$f_Y = CNN_Y(I_Y), f_{Cb} = CNN_{Cb}(I_{Cb}), f_{Cr} = CNN_{Cr}(I_{Cr}). \quad (1)$$

여기서 f_Y, f_{Cb}, f_{Cr}은 모두 2×2 픽셀 크기의 512차원 특징맵이다.

2. 주의 모듈을 이용한 특징 조합

밝기 및 색상의 지역적 패턴을 이용한 기존 방법[14]은 단순 결합을 이용하여 독립적으로 추출한 밝기 특징과 색상 특징을 조합하였다. 단순 결합은 특징별 중요도를 동등하게 고려하여 특징들을 조합하는데, 본 논문에서는 단순 결합된 특징맵에 주의 모듈을 적용하여 각 특징에 대한 중요도를 학습을 통해 도출함으로써 밝기 특징과 색상 특징 간의 관계를 적응적으로 고려하여 특징들을 조합할 수 있도록 하였다. 본 논문에서 사용된 주의 모듈(Squeeze and Excitation Block, SE Block)[11]은 그림 2에서 볼 수 있듯이 먼저 전역 평균 풀링(Global Average Pooling)을 사용하여 입력된 특징맵의 채널(Channel)별 중요 정보를 압축한다. 그 후, 완전 연결 계층(Fully Connected Layer)과 시그모이드(Sigmoid) 함수 및 ReLU 함수를 이용하여 압축된 정보로부터 채널 간의 상호작용을 학습한 뒤 채널별 중요도를 계산하고, 이에 따라 각 채널에 새로운 가중치를 부여한다.

따라서 본 논문에서는 추출된 밝기 특징과 색상 특징을 효과적으로 조합하기 위하여 먼저 특징들을 단순 결합한 후에 그림 2의 주의 모듈을 적용하여 학습을 통해 결합된 특징맵의 채널별 중요도를 도출하였다. 그리고 도출된 채널별 중요도를 바탕으로 해당 채널에 곱해질 가중치를 부여하여 결합된 특징맵을 적응적으로 재조정해주었다.

먼저, 세 개의 세부 합성곱 신경망을 이용하여 추출한 특징맵 f_Y, f_{Cb}, f_{Cr}을 단순 결합하여 특징맵 f_{cat}을 얻는다.

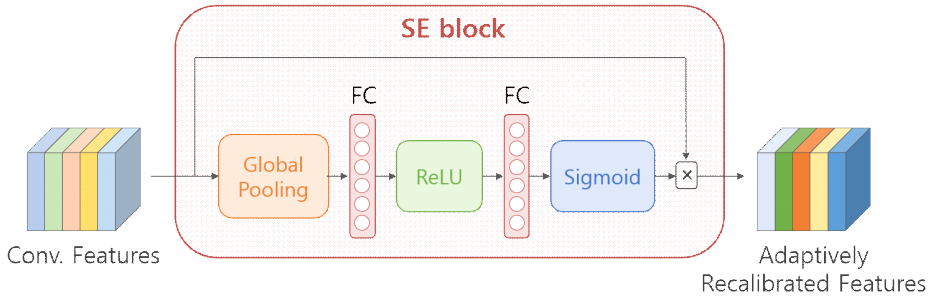

그림 2. 본 논문에서 사용된 주의 모듈의 구조

Fig. 2. Details of the attention module architecture used in this paper

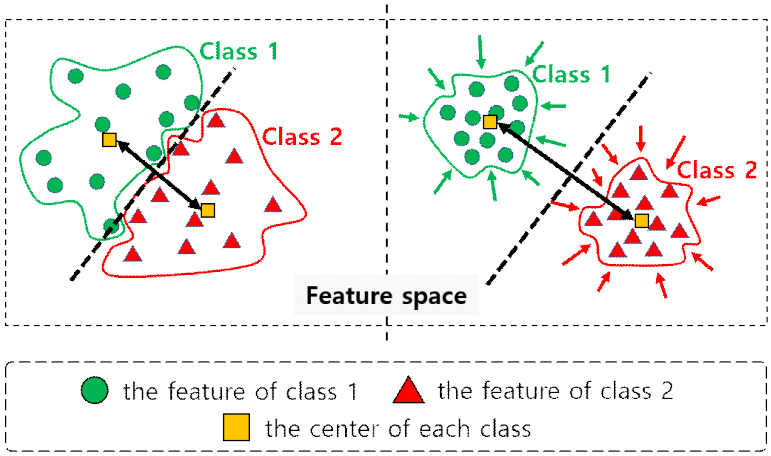

그림 3. 대비 손실함수를 통한 동일 클래스 내에 있는 특징 간의 차이 최소화 및 서로 다른 클래스에 있는 특징 간의 차이 최대화하는 과정
Fig. 3. Minmizing the intra-class features distance and maximizing the inter-class features distance by using the contrast loss

이러한 과정을 아래와 같이 표현할 수 있다.

$$f_{cat} = Concat(f_Y, f_{Cb}, f_{Cr}), \tag{2}$$

이때 $Concat(\bullet)$은 단순 결합을 의미한다. 그리고 주의 모듈을 이용하여 특징맵 f_{cat}으로부터 적응적으로 재조정된 특징맵 f을 얻는데, 이와 같은 과정은 다음과 같이 표현될 수 있다.

$$f = SE(f_{cat}), \tag{3}$$

여기서 $SE(\bullet)$은 주의 모듈을 의미한다. 3장의 실험 결과 및 분석에서 주의 모듈을 통해 적응적으로 재조정된 특징맵을 이용하는 것이 단순 결합된 특징맵을 이용하는 것보다 얼굴 위변조 여부를 판별하는데 더욱 효과적임을 보인다.

3. 분류기 학습을 위한 손실함수 설계

최종적으로 손실함수와 재조합된 특징을 이용하여 이진 분류기(Binary Classifer)를 학습하고 생성된 모델을 바탕으로 얼굴 위변조 검출을 수행한다. 일반적으로 두 가지 클래스의 데이터를 분류하는 이진 문제(Binary Problem)를 해결하기 위하여 주로 소프트맥스 손실함수(Softmax Loss Function)를 사용하지만, 본 논문에서는 기존의 소프트맥스 손실함수와 제안하는 대비 손실함수를 함께 사용하여 특징의 분별력을 향상시켜 효과적인 분류기 학습이 가능하도록 하였다. 대비 손실함수는 그림 3과 같이 특징 공간(Feature Space)상에서 동일 클래스 내에 있는 특징 간의 차이(Intra-class Distance)는 최소화 시키고 동시에 서로 다른 클래스 내에 있는 특징 간의 차이(Inter-class Distance)는 최대화 시킴으로써 특징의 분별력을 높인다.

대비 손실함수를 정의하기 위하여 먼저, 각 클래스, 즉 실제 얼굴 클래스와 위변조된 얼굴 클래스에 속하는 특징 벡터의 중심(Center Vector of the Class)을 아래와 같이 계산한다.

$$\mathbf{c}_i = \frac{1}{N_i}\sum_{k=1}^{N_i}(\mathbf{x}_k)_i, \quad i = 1, 2, \tag{4}$$

이때 $\mathbf{c}_i$는 i번째 클래스의 중심 벡터이며 첫 번째 클래스($i = 1$)와 두 번째 클래스($i = 2$)는 각각 실제 얼굴 클래스와 위변조된 얼굴 클래스를 나타낸다. 또, N_i는 i번째 클래스 내의 특징 벡터의 개수, 그리고 $(\mathbf{x}_k)_i$는 i번째 클래스의 k번째 특징 벡터를 의미한다. 또한, 클래스의 중심 벡터는 전체 학습 데이터에 대하여 계산이 되는 것이 아니라, 각

미니배치에 대해 계산이 되므로 전체 특징 벡터의 개수 N_1+N_2는 미니배치의 크기와 같다. 식 (4)를 통해 계산된 클래스 중심 벡터를 이용하여 동일 클래스 내에 있는 특징 벡터 간의 차이(D_{intra})는 각 특징 벡터와 해당 클래스 중심 벡터와의 유클라디안 거리(Euclidean Distance) 평균으로 정의한다. 또, 서로 다른 클래스에 있는 특징 벡터 간의 차이(D_{inter})는 두 클래스 중심 벡터 간의 유클라디안 거리로 정의한다. 이를 다음과 같이 표현할 수 있다.

$$D_{intra}=\frac{1}{N_1+N_2}\sum_{i=1}^{2}\sum_{k=1}^{N_i}\|(\mathbf{x}_k)_i-\mathbf{c}_i\|, \tag{5}$$

$$D_{inter}=\|\mathbf{c}_1-\mathbf{c}_2\|. \tag{6}$$

위의 식 (5)과 식 (6)를 통해 계산한 D_{intra}와 D_{inter}를 이용하여 대비 손실함수를 다음과 같이 정의한다.

$$L_C=\frac{D_{intra}}{D_{inter}}=\frac{\frac{1}{N_1+N_2}\sum_{i=1}^{2}\sum_{k=1}^{N_i}\|(\mathbf{x}_k)_i-\mathbf{c}_i\|}{\|\mathbf{c}_1-\mathbf{c}_2\|}. \tag{7}$$

식 (7)의 대비 손실함수 L_C를 이용하여 D_{intra}를 최소화 시키고, D_{inter}는 최대화 시킴으로써 결과적으로 더욱 높은 분별력을 가진 특징을 학습할 수 있다.

소프트맥스 손실함수와 대비 손실함수 계산을 위하여 먼저 식 (3)을 통해 재조합된 특징맵 f를 표 2와 같은 두 개의 완전연결계층으로 이루어진 분류기에 통과시켜 1024차원의 특징 벡터 $\mathbf{x}_f$를 얻는다. 이 특징 벡터를 이용하여 주 손실함수 L_{MAIN}를 다음과 같이 계산한다.

$$L_{MAIN}=L_S(\mathbf{x}_f)+\lambda_C L_C(\mathbf{x}_f), \tag{8}$$

이때, $L_S(\mathbf{x}_f)$와 $L_C(\mathbf{x}_f)$는 각각 특징 벡터 $\mathbf{x}_f$를 이용한 소프트맥스 손실함수와 대비 손실함수를 의미하고, λ_C은 대비 손실함수에 대한 가중치 값을 의미한다. 또한, 본 논문에서는 추출된 특징의 분별력을 효과적으로 향상시키기 위해 각 세부 합성곱 신경망에 표 2와 같은 두 개의 완전연결계층으로 이루어진 보조 신경망(Auxiliary Network)를 추가하였다. 그리고 얼굴 영역의 Y, Cb, Cr 영상으로부터 추출한 특징맵 f_Y, f_{Cb}, f_{Cr}를 각각의 보조 신경망에 통과시켜 1024차원의 특징 벡터 $\mathbf{x}_{f_Y}$, $\mathbf{x}_{f_{Cb}}$, $\mathbf{x}_{f_{Cr}}$을 얻고, 각 특징 벡터에 대한 소프트맥스 손실함수를 계산하여 이를 전체 신경망 학습을 위한 보조 손실함수로 사용하였다. 이와 같은 과정을 다음과 같이 나타낼 수 있다.

$$L_{AUX}=L_S(\mathbf{x}_{f_Y})+L_S(\mathbf{x}_{f_{Cb}})+L_S(\mathbf{x}_{f_{Cr}}), \tag{9}$$

여기서 L_{AUX}는 보조 손실함수를 의미하고, $L_S(\bullet)$는 각 특징 벡터를 이용하여 계산한 소프트맥스 손실함수를 의미한다. 따라서, 제안하는 신경망 모델을 학습시키기 위한 전체 손실함수 L는 아래와 같이 주 손실함수 L_{MAIN}과 보조 손실함수 L_{AUX}의 합으로 나타낼 수 있다.

$$L=L_{MAIN}+\lambda_{AUX}L_{AUX}, \tag{10}$$

여기서 λ_{AUX}는 보조 손실함수에 대한 가중치 값을 의미한다. 3장의 실험 결과 및 분석을 통해 제안하는 손실함수를 이용했을 때 위변조 검출 성능이 향상됨을 보인다.

표 2. 본 논문에서 사용된 보조 신경망 및 분류기의 구조. fc(N)은 N개의 뉴런을 가진 완전연결계층을 의미한다.
Table 2. Details of the classifier and the auxiliary networks used in this paper. fc(N) is a fully connected layer with N neurons.

	Auxiliary Network for CNN_Y	Auxiliary Network for CNN_{Cb}	Auxiliary Network for CNN_{Cr}	Classifier
Input	3×3×512	3×3×512	3×3×512	3×3×1536
FC-1	fc(2048)	fc(2048)	fc(2048)	fc(2048)
FC-2	fc(1024)	fc(1024)	fc(1024)	fc(1024)

Ⅲ. 실험 결과 및 분석

본 논문에서 제안하는 신경망 구조의 성능 평가를 위해 CASIA Face Anti-Spoofing Database(CASIA FASD)[17]를 이용한 얼굴 위변조 검출 실험을 수행하였다. CASIA FASD는 50명의 사용자로부터 획득된 실제 얼굴 동영상과 위변조된 얼굴 동영상으로 구성되어 있다. 실제 얼굴 동영상의 경우 저화질, 중간화질, 고화질 등 세 종류 화질의 동영상으로 이루어져 있다. 또, 위변조된 얼굴 동영상의 경우 사진 위변조 공격, 마스크 위변조 공격, 동영상 위변조 공격 등 세 가지 종류의 위변조 공격 동영상으로 이루어져 있고, 각각의 위변조 공격 동영상은 실제 얼굴 동영상과 마찬가지로 세 종류 화질의 동영상으로 구성되어 있다. 50명의 사용자 동영상 중에서 20명의 동영상은 학습을 위해 사용되었고 30명의 동영상은 성능 평가를 위해 사용되었다. 또한, 실제 얼굴 영상과 위변조된 얼굴 영상의 개수를 동일하게 맞춰주기 위하여 실제 얼굴 동영상으로부터는 무작위로 30 프레임의 영상을 추출하였고, 위변조된 얼굴 동영상으로부터는 무작위로 10 프레임의 영상을 추출하였다. 따라서, 최종적으로 1,800장의 실제 얼굴 영상과 1,800장의 위변조된 얼굴 영상이 학습에 이용되었고, 2,700장의 실제 얼굴 영상과 2,700장의 위변조된 얼굴 영상이 성능 평가에 사용되었다. 학습과 성능 평가를 위하여 Geforce GTX TITAN X가 이용되었고, 실험을 위한 코드는 파이토치(Pytorch)를 이용하여 구현하였다.

실험을 위해 원본 영상에서 얼굴 검출 알고리즘[12]을 이용하여 얼굴 영역을 검출하고, 검출된 영역을 112×112 픽셀 크기의 영상으로 재조정하였다. 이와 같이 재조정된 얼굴 영역을 각각 Y, Cb, Cr 영상으로 변환하여 각 세부 합성곱 신경망의 입력으로 사용하여 신경망 학습 및 성능 평가를 수행하였다. 학습이 완료된 후에는 시험 영상 5,400장에 대한 동일 오류율 (Equal Error Rate, EER)을 계산하여 주의 모듈의 적용 여부와 손실함수의 가중치에 따른 제안하는 신경망 구조의 성능을 비교하였다.

먼저, 밝기 정보와 색상 정보를 함께 이용한 합성곱 신경망 구조의 우수성을 보이기 위하여, Y 영상만을 입력으로 사용할 때와 Y, Cb, Cr 영상을 입력으로 사용할 때의 제안하는 신경망 구조의 성능을 비교하였다. 이때 대비 손실함수 L_C에 대한 가중치 λ_C는 0.1로 설정하였고, 보조 손실함수 L_{AUX}에 가중치 λ_{AUX}는 0.5로 설정하여 학습을 진행하였다. 표 3을 보면 Y, Cb, Cr 영상 모두를 사용했을 때가 Y 영상만을 사용했을 때 대비 동일 오류율이 낮음을 확인할 수 있다. 또한, 보조 손실함수 L_{AUX}에 대한 최적의 가중치 λ_{AUX}를 얻기 위해 λ_{AUX}을 0, 0.25, 0.5, 0.75 등 네 가지 값으로 조절하며 학습을 진행하였다. 이때, 주의 모듈이 적용되지 않은 신경망 구조가 이용되었으며, 대비 손실함수 L_C에 대한 가중치 λ_C는 0으로 설정하였다. 표 4를 보면 λ_{AUX}이 0.5일 때 가장 낮은 오류율을 보여 이후의 모든 실험은 λ_{AUX}을 0.5로 고정하여 수행하였다.

표 3. 입력 영상에 따른 제안하는 합성곱 신경망 구조의 EER 비교
Table 3. EERs comparison of the proposed convolutional neural network according to the input image

Methods	Equal Error Rate(EER) @$\lambda_C=0.1$, $\lambda_{AUX}=0.5$
Y	6.44%
YCbCr	2.93%

최종적으로 주의 모듈을 이용한 특징맵의 적응적 재조정 및 제안하는 대비 손실함수의 우수성을 보이기 위하여, 주의 모듈을 적용한 신경망 구조와 주의 모듈을 적용하지 않은 신경망 구조에 대하여 각각 실험을 수행하였다. 이때,대비 손실함수 L_C에 대한 가중치 λ_C를 0, 0.01, 0.1 등 세 가지 값으로 조절하며 학습을 진행하였다. 표 5를 보면 실험한 모든 대비 손실함수 가중치 값에 대하여 주의 모듈을

표 4. 보조 손실함수 가중치 값에 따른 제안하는 합성곱 신경망 구조의 EER 비교
Table 4. EERs comparison of the proposed convolutional neural network according to the auxiliary loss weight value

Methods	Equal Error Rate(EER) @$\lambda_C=0$			
	$\lambda_{AUX}=0$	$\lambda_{AUX}=0.25$	$\lambda_{AUX}=0.5$	$\lambda_{AUX}=0.75$
Without SE block	4.56%	4.48%	4.04%	4.15%

표 5. 주의 모듈 적용 여부와 대비 손실함수 가중치 값에 따른 제안하는 합성곱 신경망 구조의 EER 비교
Table 5. EERs comparison of the proposed convolutional neural network according to the contrast loss weight value and attention module

Methods	Equal Error Rate(EER) @$\lambda_{AUX}=0.5$		
	$\lambda_C=0$	$\lambda_C=0.01$	$\lambda_C=0.1$
Without SE block	4.04%	3.92%	3.33%
With SE block	3.74%	3.56%	2.93%

적용하였을 때가 주의 모듈을 적용하지 않았을 때보다 동일 오류율이 낮음을 알 수 있다. 또, 대비 손실함수를 추가함으로써 동일 오류율이 낮아짐을 확인할 수 있다. 표 6에서는 기존의 다양한 얼굴 위변조 검출 방법과 본 논문에서 제안하는 방법의 성능을 비교하였는데, 제안하는 방법이 기존의 방법 대비 동일 오류율을 효과적으로 개선할 수 있음을 확인하였다. 그림 4에는 제안하는 방법을 통해 얼굴 위변조 검출을 수행하였을 때 오검출된 얼굴 영상의 예를 나타내었다. 이처럼 다양한 실험을 통해 제안하는 합성곱 신경망 구조 및 손실함수가 얼굴 위변조 검출에 효과적임을 알 수 있다.

그림 4. 위: 실제 얼굴을 위변조된 얼굴이라 판단한 예시. 아래: 위변조된 얼굴을 실제 얼굴이라 판단한 예시
Fig. 4. Top: live faces detected as fake. Bottom: fake faces detected as live

Ⅳ. 결 론

본 논문에서는 얼굴 영상의 밝기 정보와 색상 정보를 함께 고려하는 합성곱 신경망 기반의 얼굴 위변조 검출 방법을 제안하였다. 기존 방법과는 달리, 주의 모듈을 적용하여 특징 간의 상관관계를 학습을 통해 도출함으로써 추출된 특징을 적응적으로 조합할 수 있도록 하였다. 또한, 동일 클래스 내 특징 간의 차이는 최소화 시키고, 동시에 서로 다른 클래스 내 특징 간의 차이는 최대화 시켜 특징의 분별력을 높일 수 있는 대비 손실함수를 제안함으로써 효율적인 분류기 학습이 가능하도록 하였다. 다양한 실험을 통해 제안하는 방법이 기존의 얼굴 위변조 검출 방법 대비 향상된 성능을 보임을 확인하였다.

참 고 문 헌 (References)

[1] A. K. Jain, K. Nandakumar, and A. Ross, “50 years of biometric research: Accomplishments, challenges, and opportunities,” Pattern Recognit. Lett., vol. 79, pp. 80 - 105, Aug. 2015.
[2] J.-W. Li, “Eye blink detection based on multiple Gabor response waves,” in Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC), Jul. 2008, pp. 2852 - 2856.
[3] G. Chetty and M. Wagner, “Multi-level liveness verification for face-voice biometric authentication,” in Proc. Biometrics Symp., Special Session Res. Biometric Consortium Conf., Sep. 2006, pp. 1 - 6.
[4] K. Kollreider, H. Fronthaler, and J. Bigun, “Evaluating liveness by face images and the structure tensor,” in Proc. 4th IEEE Workshop Automat. Identificat. Adv. Technol., Oct. 2005, pp. 75 - 80.
[5] J. Galbally, S. Marcel, and J. Fierrez, “Image quality assessment for

표 6. 제안하는 방법과 기존의 얼굴 위변조 검출 방법 간의 EER 비교
Table 6. EERs comparison between the proposed method and previous face anti-spoofing methods

Reference	Year	Methodology	EER(%)
Yan et.al.	2014	CNN + SVM [9]	7.4
Boulkenafet et.al.	2015	Color texture analysis based [13]	6.2
Xu et.al.	2015	LSTM-CNN architecture based [14]	5.1
Zhao et.al.	2017	Dynamic texture based [15]	6.5
Gan et.al.	2017	3D-CNN based [16]	6.4
Proposed Method	2019	Based on combination of luminance and chrominance with CNN	2.9

fake biometric detection: application to iris, fingerprint, and face recognition," IEEE Trans. Image Process. vol. 23, no. 2, pp. 710 – 724, Feb. 2014.

[6] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971 – 987, Jul. 2002.

[7] J. Maatta, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in Proc. Int. Joint Conf. Biometrics (IJCB), Oct. 2011, pp. 1 – 7.

[8] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, "An investigation of local descriptors for biometric spoofing detection," IEEE Trans. Inf. Forensics Security, vol. 10, no. 4, pp. 849 – 863, Apr. 2015.

[9] J. Yang, Z. Lei, and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," arXiv preprint arXiv: 1408.5601, Aug. 2014.

[10] O. Lucena, A. Junior, V. Moia, R. Souza, E. Valle, and R. Lotufo, "Transfer learning using convolutional neural networks for face anti-spoofing," in Proc. Int. Conf. Image Anal. Recognit., Jun. 2017, pp. 27 – 34.

[11] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), June. 2018, pp. 7132-7141.

[12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), Dec. 2001, pp. I-511 – I-518.

[13] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing using color texture analysis," IEEE Trans. Inf. Forensics Security, vol. 11 no. 8, pp. 1818 – 1830, Aug. 2016.

[14] Z. Xu, S. Li, and W. Deng, "Learning temporal features using lstm-cnn architecture for face anti-spoofing," in Proc. IAPR Asian Conf. Pattern Recognit. (ACPR), Nov. 2015, pp. 141-145.

[15] X. Zhao, Y. Lin, and J. Heikkilä, "Dynamic texture recognition using volume local binary count patterns with an application to 2D face spoofing detection," IEEE Trans. Multimedia, vol. 20, no. 3, pp. 552-566, Mar. 2017

[16] J. Gan, S. Li, Y. Zhai, and C. Liu, "3D convolutional neural network based on face anti-spoofing," in Proc. Int. Conf. Multimedia Image Process. (ICMIP), Mar. 2017, pp. 1-5.

[17] Z. Zhang, J. Yan, S. Liu, Z. Lei, D Yi, S. Z. Li. "A Face Antispoofing Database with Diverse Attacks." in Proc. Int. Conf. Biometrics (ICB), Mar. 2012, pp. 26 – 31

저 자 소 개

김 은 석

- 2018년 2월 : 건국대학교 학사
- 2018년 3월 ~ 현재: 건국대학교 전기전자공학부 석사과정
- ORCID : https://orcid.org/0000-0003-4818-2221
- 주관심분야 : 컴퓨터 비전, 생체인식, 기계학습, 패턴 인식

김 원 준

- 2012년 8월 : 한국과학기술원(KAIST) 박사
- 2012년 9월 ~ 2016년 2월 : 삼성종합기술원 전문연구원
- 2016년 3월 ~ 현재 : 건국대학교 전기전자공학부 조교수
- ORCID : https://orcid.org/0000-0001-5121-5931
- 주관심분야 : 영상이해, 컴퓨터 비전, 기계학습, 패턴 인식