JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 27 No. 9, pp. 41-47, September 2022
https://doi.org/10.9708/jksci.2022.27.09.041

Facial Expression Recognition through Self-supervised Learning for Predicting Face Image Sequence

Yeo-Chan Yoon*, Soo Kyun Kim**

*Professor, Dept. of Artificial Intelligence, Jeju National University, Jeju, Korea
**Professor, Dept. of Computer Engineering, Jeju National University, Jeju, Korea

[Abstract]

In this paper, we propose a new and simple self-supervised learning method that predicts the middle image of a face image sequence for automatic expression recognition. Automatic facial expression recognition can achieve high performance through deep learning methods, however, generally requires a expensive large data set. The size of the data set and the performance of the algorithm are tend to be proportional. The proposed method learns latent deep representation of a face through self-supervised learning using an existing dataset without constructing an additional dataset. Then it transfers the learned parameter to new facial expression reorganization model for improving the performance of automatic expression recognition. The proposed method showed high performance improvement for two datasets, CK+ and AFEW 8.0, and showed that the proposed method can achieve a great effect.

▸ **Key words**: Facial Expression Recognition, Unsupervised-learning, Self-supervised Learning, Deep Learning, Artificial Intelligence

[요 약]

본 논문에서는 자동표정인식을 위하여 얼굴 이미지 배열의 가운데 이미지를 예측하는 새롭고 간단한 자기주도학습 방법을 제안한다. 자동표정인식은 딥러닝 모델을 통해 높은 성능을 달성할 수 있으나 일반적으로 큰 비용과 시간이 투자된 대용량의 데이터 세트가 필요하고, 데이터 세트의 크기와 알고리즘의 성능이 비례한다. 제안하는 방법은 추가적인 데이터 세트 구축 없이 기존의 데이터 세트를 활용하여 자기주도학습을 통해 얼굴의 잠재적인 심층표현방법을 학습하고 학습된 파라미터를 전이시켜 자동표정인식의 성능을 향상한다. 제안한 방법은 CK+와 AFEW 8.0 두 가지 데이터 세트에 대하여 높은 성능 향상을 보여주었고, 간단한 방법으로 큰 효과를 얻을 수 있음을 보여주었다.

▸ **주제어**: 얼굴인식, 비교사학습, 자기주도학습, 딥러닝, 인공지능

• First Author: Yeo-Chan Yoon, Corresponding Author: Soo Kyun Kim
*Yeo-Chan Yoon (ycyoon@jejunu.ac.kr), Dept. of Artificial Intelligence, Jeju National University
**Soo Kyun Kim (kimsk@jejunu.ac.kr), Dept. of Computer Engineering, Jeju National University
• Received: 2022. 08. 02, Revised: 2022. 09. 05, Accepted: 2022. 09. 06.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

인공지능 기술을 활용한 자동표정인식 기술은 휴먼컴퓨팅인터페이스(HCI)의 핵심기술 중 하나로, 보안, 로보틱스, 의료, 교육, 엔터테인먼트 등 다양한 분야에서 활용되고 있다. 표정은 인간이 감정을 표현하는 데 가장 많이 활용되는 모달리티(Modality)로 상대방의 표정을 통해 기분과 상태를 분석하여 적절히 대응하는 시스템을 제작할 수 있다. 최근에는 딥러닝 기술을 적용하여 표정인식 기술을 고도화하는 다양한 시도[1]가 이루어지고 있다.

그러나, 딥러닝 기술을 활용하여 성공적으로 표정을 통해 감성을 인식하기 위해서는 매우 많은 양의 레이블 데이터가 필요하다. 레이블 데이터를 구축하기 위해서는 필수적으로 훈련받은 사람이 직접 태깅해야 하며 이는 많은 시간과 비용이 소모된다. 따라서 전이학습(Transfer Learning), 자기지도학습(Self-supervised Learning) 등 비교사 학습방법(Unsupervised Learning)이 최근에 주목을 받고 있다. 비교사 학습방법은 사람이 직접 태깅한 데이터 없이 손쉽게 얻을 수 있는 원시(raw)데이터만을 이용하여 학습하는 기계학습 방법을 지칭한다. 비교사 학습방법을 이용하면 잠재적인 이미지의 표현방식을 미리 학습하고 이를 이용하여 성능을 향상할 수 있다. 자기지도학습은 비교사 학습 방법의 하나로, 사람의 개입 없이 자동으로 정확하게 레이블을 생성하여 학습데이터를 구축하여 잠재적인 표현방식을 학습하는 방법이다.

최근에 자기주도학습을 통해 이미지 인식의 성능을 향상한 많은 연구가 소개되었다. Zhang[2]과 Larsson[3]은 컬러이미지를 흑백으로 변환시켜 학습데이터를 자동으로 구축한 후에 이를 다시 컬러 채색하도록 CNN을 학습시키는 방식의 자기주도학습을 제안하였다. Doersch[4]와 Noroozi[5]는 하나의 사진을 여러 개의 패치로 나눈 후에 패치의 위치를 예측하는 방식의 자기주도학습을 제안하였으며, Chun-Liang[6]은 이미지 기반 이상탐지에 활용하기 위하여 이미지의 작은 패치를 다른 패치로 교체하는 방식의 자기주도학습 방법을 제안하였다.

본 논문은 자동표정인식 기술에 특화된 새로운 자기지도학습 방법을 제안하고, 이를 이용하여 높은 성능의 자동표정인식 모델을 구축하는 것을 목표로 한다. 본 논문에서 제안하는 프레임워크는 표정인식을 위해 2단계 파이프라인으로 구성되어 있다. 첫 번째 단계에서는 자기주도학습을 이용하여 표정인식을 위한 잠재적인 심층표현을 학습한다. 제안하는 자기주도학습은 연속된 표정이미지를 입력으로 하여 중간단계의 표정을 인식하도록 학습한다. 중간

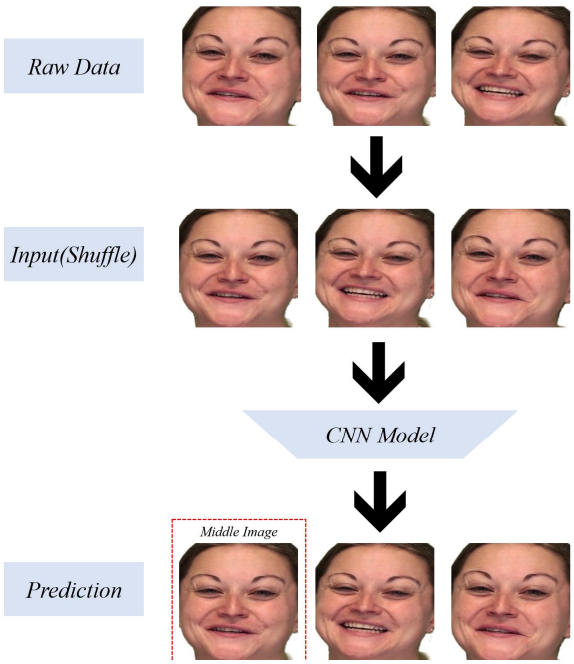

Fig. 1. The proposed self-supervised learning method for predicting the middle image in an image sequence.

표정을 학습하기 위해서는 이미지에서 표정을 표현하기 위한 잠재적인 심층 표현방법을 익혀야 가능하다. 이러한 직관에 따라 자기주도학습을 통해서 표정을 표현하는 눈, 얼굴 등의 주요 기관의 움직임을 분석할 수 있다. 그림1은 제안하는 중간표정 예측을 통한 자기주도학습을 보여준다. 원시데이터의 프레임 순서를 랜덤하게 섞은 후에 중간 이미지를 예측하는 방식을 통해서 얼굴 이미지에 대한 잠재적 심층 표현방법을 학습한다.

두 번째로 자기주도학습을 통해 학습한 모델을 파인튜닝(finegrained-tuning)하여 자동표정인식을 수행한다. 첫 번째 단계에서 학습한 잠재심층표현방식을 활용하여 정확한 감성을 인식할 수 있도록 하며, 이를 위하여 신경망 모델에서 마지막 분류를 위한 선형 계층을 제외한 다른 모든 계층을 파인튜닝에 활용하여 표정을 통한 사람의 감성을 인식한다.

이 논문의 공헌점은 다음과 같다.

- 본 논문에서는 표정인식을 위한 새롭고 간단하며, 효과적인 자기주도학습 방법을 제안한다. 제안하는 방법은 얼굴에 대한 잠재 심층 표현방식을 성공적으로 학습한다.

- 제안한 방법을 표정인식 도메인에 적용하고 두 가지 주요 표정인식 데이터 세트에 실험하여 중간 얼굴 이미지 예측 자기주도학습을 통한 자동표정인식 연구가 효과적임을 입증한다.

II. Related works

2.1 Facial Expression Recognition

Ekman[7]은 다양한 문화를 분석하고 이를 기반으로 사람이 자각하는 기본적인 감성을 6가지로 분류하였다. 분류된 감성은 분노, 역겨움, 공포, 행복, 슬픔, 놀라움이었다. 이에 기반하여 다양한 연구들이 자동으로 감정을 분류하려고 시도하였고, 6가지 감정 이외에도 중립감정, 증오 등의 감정을 더 해 자동표정인식의 효용성을 높이고 성능을 향상했다. 표정을 통한 감성인식은 정적인 하나의 이미지를 인식하는 방법[8]과 비디오와 같이 연속된 이미지를 입력으로 받아서 표정기반의 감정인식을 하는 연구[1], [9]로 나눌 수 있다. 최근에는 셀프어텐션(Self-attention)기술을 활용하여 연속된 이미지 간의 관계를 분석하고 효과적으로 표정을 인식하는 연구가 보고된다. FAN[9]은 비디오 기반의 자동표정인식을 위해 셀프어텐션을 사용한 네트워크이다. 여러 장의 얼굴 이미지를 입력으로 받아서 표정인식을 위한 고정된 크기의 자질 벡터를 생성한 후에 프레임 어텐션 모듈을 통해 프레임 자질 벡터 간의 관계를 분석하여 중요한 자질벡터에 최종 자동표정인식 결과 도출을 위하여 더 많은 가중치를 줄 수 있도록 학습한다. Li[10]는 조사 논문을 통해 딥러닝 기술을 활용한 다양한 얼굴인식 기반 감정분류 기술에 대하여 세부적으로 기술하였다.

본 논문에서는 연속된 이미지 기반의 표정인식 기술을 연구한다. 효과적으로 연속된 이미지 기반의 자동표정인식 모델을 학습시키기 위하여 어텐션 기반의 CNN 모델인 FAN[9]을 백본(Back-bone)으로 사용하고 자기주도 학습을 이용하여 성능을 개선한다.

2.2 Self-supervised learning

자기주도학습은 고비용의 레이블 데이터 구축을 지양하는 비교사학습방법의 일종으로, 사람의 개입 없이 레이블을 자동생성하여 학습에 활용하는 기술이다. 이는 전이학습(Transfer learning)[11]의 일종으로 볼 수 있는데, 전이학습은 다른 태스크 혹은 도메인을 위해 구축된 대량의 데이터를 이용하여 잠재 심층표현을 사전에 학습하고 이를 목적 태스크 혹은 도메인에 사용하여 성능을 향상하는 기술이다.

최근 몇 년간 목적 태스크의 성능을 향상하기 위하여 다양한 자기주도학습이 연구되었다. Zhou[12]는 인간의 감정을 예측하기 위하여 다른 센서로부터 습득 가능한 데이터를 활용하여 자기주도학습에 사용하였다. Doersch[4]는 두 이미지 패치의 상대적인 위치를 분류하기 위하여 CNN을 이용하였다. 이미지 패치를 지우거나 복사, 붙여넣기를 하여 이미지 자기주도학습을 시도한 연구도 보고된다. 이미지를 잘라내고[13] 랜덤하게 이미지의 특정 부분을 잘라내어[14] CNN의 안정성을 높이고, 이를 발전시켜 길고 가느다란 사각형을 랜덤한 색으로 잘라내고 붙여내어 성능을 개선했다[15]. Gidaris[16]는 이미지를 회전시킨 후에, 회전 각도를 예측하는 자기주도학습을 제안하고 이를 통하여 객체인식기술의 성능을 향상했다.

본 논문에서는 얼굴표정인식을 위하여 얼굴의 변화를 자기주도학습을 통하여 분석하고 이를 활용할 수 있는 새로운 방법을 제안한다.

III. The Proposed Method

3.1 The proposed self-supervised learning method with central image prediction

본 논문의 목표는 비교사 방식으로 인간의 개입 없이 자동으로 구축한 학습데이터 세트를 이용하여 표정을 표현하는 잠재심층표현을 학습하고 이를 활용하여 전이학습을 통해 목표로 하는 자동표정인식 기술의 성능을 높이는 데 있다. 이를 위하여 본 논문에서는 CNN 기반의 딥러닝 모델 F를 연속된 이미지 중 가운데 이미지를 예측할 수 있도록 학습한다. 함수 g는 원본 이미지 배열 X에서 중간 이미지를 y번째 이미지로 이동시키고 나머지 이미지들은 랜덤하게 섞는 함수로 X를 입력으로 받아 가운데 이미지가 k번째에 위치하고 나머지가 랜덤하게 배치된 X^{y*}를 출력한다. 함수를 활용하여 자동으로 생성한 데이터를 자기주도학습을 위한 학습 셋으로 사용할 수 있다. y는 하나의 비디오에 구성된 이미지의 개수가 n일 때 1에서 n까지의 값을 가진다. 제안하는 방법은 아래 수식과 같이 정의된다.

$$F(y=k \mid \theta, X^{y^*})_{k=1}^{n} = F(y=k \mid \theta, g(X, y=k))_{k=1}^{n} \quad (1)$$

수식(1)은 학습 가능한 파라미터 Θ가 주어졌을 때 이미지 배열 X를 함수 g를 이용하여 중간 이미지가 k번째에 위치하게 하고 나머지는 랜덤하게 섞도록 한 후에 중간 이미지 위치 k를 예측하기 위한 확률 분포를 나타낸다.

따라서 N개의 학습 이미지가 주어졌을 때 자기주도학습으로 딥러닝 모델을 학습시키기 위한 목적함수는 다음 수식[16]을 만족시켜야 한다.

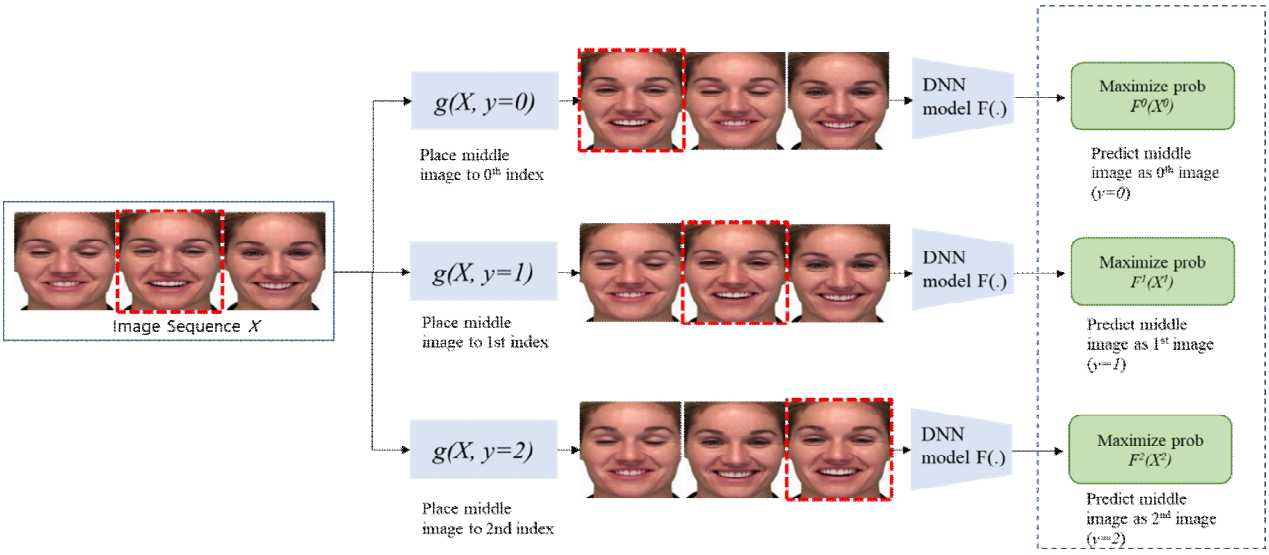

Fig. 2. System architecture of the self-supervised learning method to predict the central image.

$$\min_{\theta}\frac{1}{N}\sum_{i=1}^{N} loss(X_i,\theta), \qquad (2)$$

위 수식에서 로스 함수는 아래와 같이 정의된다.

$$loss(X_i,\theta) = -\frac{1}{n}\sum_{k=1}^{n}\log(F(y=k \mid \theta, X^{y^*})). \qquad (3)$$

확률 분포 수식(1)에 따라 크기 n의 이미지 배열 X가 주어졌을 때 가운데 이미지가 놓일 수 있는 모든 경우의 수 n에 대하여 각각 확률을 계산한 후에 평균을 내어서 로스 값을 계산한다.

그림 2는 본 논문에서 제안하는 자기주도학습 모델의 구조도이다. 3장 이상의 이미지 열(Image Sequence)을 입력으로 받아, 이미지 열의 순서를 바꾼 후에 가운데 이미지의 위치를 예측하는 방식으로 잠재심층표현을 학습한다. 제안하는 알고리즘은 미리 태깅된 학습데이터 없이 딥러닝 학습을 수행한다. 이미지 배열은 동영상에서 순서대로 이미지를 추출하는 방식으로 쉽게 획득할 수 있다. 이미지 학습데이터를 생성하는 함수 $g(X,y)$의 두 번째 인자를 0으로 설정하였을 때, g 함수는 가운데 이미지를 0번 인덱스로 옮겨서 그림 2의 최상위 학습 예제를 구성한다. 두 번째 인자를 1로 설정하였을 때 g 함수는 중간 이미지를 1번 인덱스로 옮겨서 그림2의 중간과 같이 학습 예제를 구성한다. 마지막으로 y=2일 때에는 그림 2 예제의 최하단 그림과 같이 학습 예제를 구성한다. 구성된 학습 예제를 통해서 목표 DNN(Deep Neural Network) 모델 F의 가중치(Weight)를 학습시키는데, 이때 모델 F를 이용해서 주어진 이미지 들 중 가운데에 위치할 이미지의 인덱스를 찾는 형식으로 학습한다. 이러한 자기주도학습은 전이학습을 통해 이미지 분류의 성능을 높이기 위하여 사전에 수행된다. 그림2를 통해 학습한 DNN Model F의 가중치(Weight)는 목표로 하는 이미지 분류 시스템의 모델 F'에 전이된다. 새로 학습할 모델 F'는 전이학습을 이용하여 전이학습 없이 모델을 학습할 때보다 높은 성능을 보여준다.

3.2. Transfer Learning

본 논문에서는 3.1의 자기주도학습을 통해 생성한 모델의 파라미터를 자동표정인식 모델의 초기 파라미터로 사용하는 전이학습을 이용하여 성능을 향상한다. 그림3은 본 논문에서 제안하는 전이학습을 이용하여 파라미터를 전이시키는 전이학습의 구조도를 나타낸다. 자기주도학습을 통해서 표정을 나타내는 주요 기관인 눈, 코, 입의 움직임을 분석할 수 있는 파라미터를 학습하고 이를 이용하여 자동표정인식을 효과적으로 수행할 수 있다.

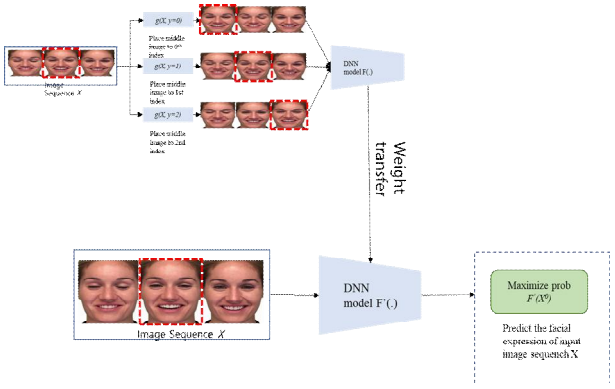

Fig. 3. Transfer learning with the self-supervised learning with central image prediction

3.3. Backbone Network

본 논문에서는 자기주도학습과 자동표정인식을 위한 백본 네트워크로 Meng[9]이 제안한 Frame Attention Networks (FAN)를 사용한다. FAN은 셀프어텐션 기반의 네트워크로 연속된 이미지 열의 변화를 분석하는데 효율적인 것으로 알려져 있다.

IV. Experimental Results

이 절에서는 제안하는 중간 이미지 예측 자기주도학습을 통한 전이학습을 자동표정인식 태스크에 도입하였을 때 성능을 분석하도록 한다. 제안하는 방법을 공개된 비디오 기반의 자동표정인식 데이터 세트 2개에 적용해보고 성능 향상 폭을 분석하고, 자기주도학습 시에 몇 개의 이미지를 사용하였을 때 효과가 높은지도 추가로 분석하도록 한다.

4.1 Dataset

자동표정인식을 위하여 2개의 공개된 데이터 세트를 활용한다. 각 데이터 세트에 대한 설명은 다음과 같다.

CK+ [17]: 자동표정인식을 위하여 실험실 환경에서 구축한 데이터베이스로 123개의 주제에 대한 593개의 비디오 배열로 구성되어 있다. 비디오는 중립이미지에서 정점 표현으로 표정이 이동되도록 촬영하였다. 각각의 비디오는 분노, 경멸, 역겨움, 공포, 행복, 슬픔, 놀람의 7가지 감정 중 하나로 태깅되었고 테스트를 위하여 일반적으로 10-폴드 크로스 밸리데이션(10-fold cross validation) 방법이 사용된다. 10-폴드 크로스 밸리데이션은 데이터 세트의 규모가 작은 경우에 실험결과에 대한 신뢰성을 높이기 위하여 사용되는 방법으로, 전체 데이터 세트를 10개로 나눈 후에 9개의 서브 셋을 학습데이터로, 1개의 서브 셋을 평가데이터 세트로 사용한다. 이때 평가데이터 세트로 사용되는 서브 셋을 교체해 가며 총 10번의 실험을 하고 평균치를 최종 성능으로 사용한다.

AFEW 8.0 [18]: 2013년부터 시작된 EmotiW 워크샵을 위한 평가 플랫폼으로 분노, 역겨움, 공포, 행복, 슬픔, 놀람, 중립의 7가지의 감성이 태깅되어 있다. 영화 및 TV 시리즈에 출현한 배우의 표정 표현을 수집하여 태깅하였고 학습셋 773개, 검증셋 383개, 테스트셋 653개로 구성되어 있다.

4.2 Facial Image Sequence Prediction

Table 1. Central image prediction performance for the CK+ set.

# of input images	Accuracy
3	91.2%
5	88.1%
7	82.3%
9	76.7%

Table 2. Central image prediction performance for the AFEW 8.0 set.

# of input images	Accuracy
3	83.5%
5	81.4%
7	76.8%
9	67.1%

테이블 1, 2는 각각 CK+ 데이터 세트와 AFEW 데이터 세트에 대하여 자기주도학습을 통한 중간 이미지 예측 성능을 보여준다. CK+의 경우 정면 이미지 위주여서 비교적 태스크 난이도가 쉬워서 AFEW에 비하여 높은 성능을 보여준다. 직관적으로 예측할 수 있듯이 입력 이미지 숫자가 많을수록 중간 이미지를 예측하기 어렵다. 3장에 대해서는 CK+ 셋에 대하여 91.2%의 성능을 보여주었지만 9장에 대해서 AFEW 8.0 데이터 세트에 대해서는 67.1%로 상대적으로 낮은 정확도를 보여주었다.

4.3. Facial Expression Recognition

테이블 3, 4는 각각 CK+ 데이터 세트와 AFEW 8.0 데이터 세트에 대하여 제안한 자기주도학습을 이용하여 파라미터를 전이하고 학습하였을 때의 자동표정인식 성능을 베이스라인과 비교하여 제시한다. 백본 네트워크로 FAN 알고리즘을 사용하였고, 제안한 자기주도학습의 효과를 입증하기 위하여 이미지 회전 각도를 예측하는 Gidaris[16]의 연구와 비교하였다.

Table 3. Facial Recognition Performance comparison on the CK+ dataset

Method	Accuracy
FAN	99.08
+Middle Image prediction Self-supervised Learning (proposed)	99.12
+Image rotation prediction Self-supervised Learning (Gidaris)	99.51

Table 4. Facial Recognition Performance comparison on the the AFEW 8.0 dataset

Method	Accuracy
FAN	50.92
+Middle Image prediction Self-supervised Learning (proposed)	51.31
+Image rotation prediction Self-supervised Learning (Gidaris)	53.17

두 개의 데이터 세트 모두에서 본 논문에서 제안한 중간 이미지 예측 자기주도학습 방법을 사용하였을 때 자동표정인식 성능이 향상된 것을 확인할 수 있다. CK+ 데이터 세트의 성능 향상도가 AFEW8.0 보다 미흡한데, 이는 기존의 성능이 워낙 높아서이다. 그러나 테이블 5와 같이 미세한 표정에 대하여 기존 방법보다 강건한 결과를 보여주는 것을 확인할 수 있다. 제안한 방법은 비교한 회전예측 기반의 기존연구보다 높은 성능을 보여주었는데 이는 이미지 배열에 대한 고려 없이 사진을 단순 회전한 것보다 표정의 미묘한 변화를 분석할 수 있는 이미지 배열을 분석한 자기주도학습이 더 효과적임을 보여준다. 테이블 5는 표정인식 결과를 보여주는 예제이다. 제안한 방법이 미묘한 표정에 대하여 다른 두 방법보다 정확히 예측하는 것을 확인할 수 있다.

그림 4과 그림5는 각각의 데이터 세트에서 자기지도학습시에 이미지 배열 크기에 따른 성능을 보여준다. CK+에서는 입력 이미지 수가 3개인 자기주도학습을 사용하였을 때 가장 높은 성능이 도출되었고, AFEW 8.0에서는 5개의 입력 이미지를 사용한 자기주도학습을 사용하였을 때 가장 높은 성능을 도출하였다. CK+의 경우 정면 이미지로 학습 셋이 구성되어 복잡도가 낮지만, AFEW 8.0은 다양한 각도에서의 얼굴 이미지로 구성되어 있어서 복잡도가 높아 상대적으로 난이도가 더 높은 5장의 이미지에 대하여 자기주도학습한 결과를 적용하였을 때 복잡한 이미지를 더 잘 처리하는 것으로 보인다.

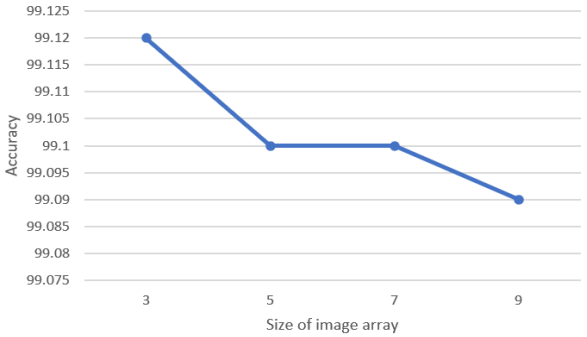

Fig. 4. Performance change according to image array size on the CK+ dataset

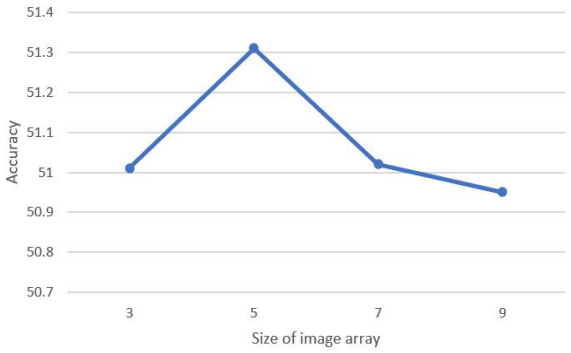

Fig. 5. Performance change according to image array size on the AFEW 8.0 dataset

V. Conclusions

본 논문에서는 얼굴 이미지 배열이 주어졌을 때 가운데 이미지를 예측하는 방식으로 자기주도학습을 하여 표정의 심층표현방법을 학습하고 이를 통하여 같은 양의 데이터 세트를 이용해서 자동표정인식의 성능을 개선하는 방법을 제안한다. 제안한 방법은 CK+와 AFEW 8.0 두 가지 데이터 세트에 대하여 높은 성능 향상을 보여주었고, 간단한 방법으로 큰 효과를 얻을 수 있음을 보여주었다. 향후 계획으로 본 논문에서 적용한 방법을 얼굴인식 등 다른 태스크에 적용할 계획이다.

Table 5. Prediciont results of each algorithm

(fear)

FAN	Proposed	Gidaris
suprise	**fear**	suprise

(happy)

FAN	Proposed	Gidaris
sadness	**happy**	**happy**

(contempt)

FAN	Proposed	Gidaris
sadness	**contempt**	disgust

ACKNOWLEDGEMENT

"This work was supported by the 2022 education, research and student guidance grant funded by Jeju National University"

REFERENCES

[1] Y. C. Yoon, "Can We Exploit All Datasets? Multimodal Emotion Recognition Using Cross-Modal Translation," in IEEE Access, vol. 10, pp. 64516-64524, 2022, doi: 10.1109/ACCESS.2022.3183587.

[2] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In European Conference on Computer Vision, pp. 649–666. Springer, 2016a.

[3] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Learning representations for automatic colorization. In European Conference on Computer Vision, pp. 577–593. Springer, 2016.

[4] Carl Doersch, Abhinav Gupta, and Alexei A Efros. Unsupervised visual representation learning by context prediction. In Proceedings of the IEEE International Conference on Computer Vision, pp. 1422–1430, 2015.

[5] Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In European Conference on Computer Vision, pp. 69–84. Springer, 2016.

[6] Li, Chun-Liang, et al. "Cutpaste: Self-supervised learning for anomaly detection and localization." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[7] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion." Journal of personality and social psychology, vol. 17, no. 2, pp. 124–129, 1971.

[8] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on. IEEE, 2016, pp. 1–10.

[9] Meng, Debin, et al. "Frame attention networks for facial expression recognition in videos." 2019 IEEE international conference on image processing (ICIP). IEEE, 2019.

[10] Li, Shan, and Weihong Deng. "Deep facial expression recognition: A survey." IEEE transactions on affective computing (2020).

[11] Weiss, Karl, Taghi M. Khoshgoftaar, and DingDing Wang. "A survey of transfer learning." Journal of Big data 3.1 (2016): 1-40.

[12] Zhou, Tinghui, et al. "Unsupervised learning of depth and ego-motion from video." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[13] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552, 2017. 2, 5

[14] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In AAAI, 2020. 5, 12

[15] Bergmann, Paul, et al. "The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection." International Journal of Computer Vision 129.4 (2021): 1038-1059.

[16] Gidaris, Spyros, Praveer Singh, and Nikos Komodakis. "Unsupervised representation learning by predicting image rotations." arXiv preprint arXiv:1803.07728 (2018).

[17] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews, "The extended cohnkanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in CVPRW, 2010.

[18] Abhinav Dhall, Amanjot Kaur, Roland Goecke, and Tom Gedeon, "Emotiw 2018: Audio-video, student engagement and group-level affect prediction," arXiv preprint:1808.07773, 2018

Authors

Yeo-Chan Yoon received BS, MS and Ph.D degrees in computer science and engineering from Korea University, Seoul, Rep. of Korea, in 2004, 2007 and 2020 respectively. Currently, he is an assistant professor in the Department of Artificial Intelligence at Jeju National University, Jeju-si, Rep. of Korea. His research interests include Deep Learning, Vision, Natural Language Processing and machine learning.

Soo Kyun Kim received Ph.D. in Computer Science & Engineering Department of Korea University, Seoul, Korea, in 2006. He joined the Telecommunication R&D Center at Samsung Electronics Co., Ltd., in 2006 and 2008. He is now a professor at the Department of Computer Engineering at Jeju National University, Korea. Dr. Kim has published many research papers in international journals and conferences. His research interests include multimedia, pattern recognition, image processing, mobile graphics, geometric modeling, and interactive computer graphics. He is a member of ACM, IEEE, IEEE CS, KACE, KMMS, KKITS, and KIIT.