JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 27 No. 6, pp. 43-52, June 2022
https://doi.org/10.9708/jksci.2022.27.06.043

Frontal Face Video Analysis for Detecting Fatigue States

Simyeong Cha*, Jongwoo Ha*, Soungwoong Yoon*, Chang-Won Ahn*

*Researcher, VAIV Company Inc., Sejong, Korea
*Researcher, VAIV Company Inc., Seoul, Korea
*Principal researcher, VAIV Company Inc., Seoul, Korea
*Director, VAIV Company Inc., Sejong, Korea

[Abstract]

We can sense somebody's feeling fatigue, which means that fatigue can be detected through sensing human biometric signals. Numerous researches for assessing fatigue are mostly focused on diagnosing the edge of disease-level fatigue. In this study, we adapt quantitative analysis approaches for estimating qualitative data, and propose video analysis models for measuring fatigue state. Proposed three deep-learning based classification models selectively include stages of video analysis: object detection, feature extraction and time-series frame analysis algorithms to evaluate each stage's effect toward dividing the state of fatigue. Using frontal face videos collected from various fatigue situations, our CNN model shows 0.67 accuracy, which means that we empirically show the video analysis models can meaningfully detect fatigue state. Also we suggest the way of model adaptation when training and validating video data for classifying fatigue.

▸ **Key words**: Fatigue measurement, Video analysis, Migration, Video classification, Deep learning, Machine learning

[요 약]

사람이 느끼는 피로는 다양한 생체신호로부터 측정이 가능한 것으로 알려져 있으며, 기존 연구는 질병과 관련된 심각한 피로수준을 산정하는데 주된 목적을 두고 있다. 본 연구에서는 피실험자의 영상을 이용하여 딥러닝 기반의 영상 분석 기술을 적용, 피로 여부를 판단하기 위한 모델을 제안한다. 특히 화상 분석에서 통상적으로 사용되는 객체 인식, 요소 추출과 함께 영상 데이터의 시계열적 특성을 고려하여 방법론을 교차한 3개 분석모델을 제시했다. 다양한 피로상황에서 수집된 정면 얼굴 영상 데이터를 이용하여 제시된 모델을 실험하였으며, CNN 모델의 경우 0.67의 정확도로 피로 상태를 분류할 수 있어 영상 분석 기반의 피로 상태 분류가 유의미하다고 판단된다. 또한 모델별 학습 및 검증 절차 분석을 통해 영상 데이터 특성에 따른 모델 적용방안을 제시했다.

▸ **주제어**: 피로도 측정, 영상 분석, 영상 분류, 딥러닝, 머신러닝

• First Author: Simyeong Cha, Corresponding Author: Jongwoo Ha
*Simyeong Cha (casim0@vaiv.kr), VAIV Company Inc.
*Jongwoo Ha (jwha@vaiv.kr), VAIV Company Inc.
*Soungwoong Yoon (swyoon@vaiv.kr), VAIV Company Inc.
*Chang-Won Ahn (ahn@vaiv.kr), VAIV Company Inc.
• Received: 2022. 05. 20, Revised: 2022. 06. 13, Accepted: 2022. 06. 13.

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

I. Introduction

복잡한 현대 사회의 특징 중 하나는 모든 사람들이 물리적 및 가상의 환경과 다양한 상호작용을 하면서 살아가고 있다는 점이며, 이러한 상황에서 피로는 필연적으로 발생한다. 피로는 사람들이 경험하는 환경과 상호작용만큼이나 원인과 형태도 다양한데, 특히 현대인의 생활과 업무에 매우 중대한 영향을 끼친다.

기존의 피로 관련 연구는 설문조사, 생화학/생리학적 검사, 반응속도 테스트 등 다양한 생체신호의 측정을 통해 피로의 정도를 판단하여 그 영향도를 질병의 관점에서 파악하는데 주된 목적이 있었다. 이와 같은 연구는 병적 피로의 원인과 정도를 특정하는 데는 용이하나, 다양한 원인과 형태로 나타나는 피로 정도를 측정하는 데는 제한된다. 한편 최근에는 머신러닝 (Machine learning) / 인공지능 (AI) 기술의 발전으로 인해 수많은 데이터를 분석하여 다양한 분야에서 활용하고자 시도되고 있으며, 특히 컴퓨터 비전 (Computer vision) 분야의 영상 분석 기술의 발전에 따라 특정 분야에서는 인간의 판단 영역까지 도달하는 분석결과를 나타내고 있다. 특히 행동분석 등 영상분석 기술은 상태 및 행동 등을 측정하여 차후 행동 또는 감성 등을 효과적으로 분석하고 있다.

본 연구에서는 피험자의 얼굴 부분을 촬영한 영상을 딥러닝 기반 영상 분석 기법에 기반한 모델로 분석하여 피실험자의 특정 상태를 추정할 수 있는 기법을 연구했다. 구체적으로 피실험자의 영상과 촬영 당시의 피로 정도를 수집한 데이터를 이용하여 영상분석을 통한 피로 수준을 분류하는 모델을 제시하였으며, 객체 인식 (Object detection)을 이용한 얼굴 검출 여부와 영상의 시계열적 요소를 이용하는 방식에 따라 3개의 분석모델을 실험했다. 실험 결과 영상 분석을 통하여 특정 상태인 피로 수준을 유의미하게 인식할 수 있었으며, 특히 객체 인식을 이용한 모델보다 이용하지 않은 모델이 보다 높은 효율을 나타냈다. 본 연구는 영상 분석을 통해 피로 등 사람의 활동에 영향을 미치는 요소에 대한 판단을 보조할 수 있음을 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 기존에 진행된 피로 및 영상분석 기술 연구를 검토한다. 3장에서는 제안하는 영상분석 기반 모델들을 기술하고, 4장에서는 제안한 모델들의 실험 결과를 기술한다. 5장에서는 결론 및 향후 연구계획을 제시한다.

II. Related works

1. Measuring Fatigue and Fatigue Levels

피로는 생리적 특성, 수면장애, 개인의 생활 특성, 스트레스, 생활환경, 건강 유지를 위한 활동 등이 그 원인으로 알려져 있다. 이 중 수면장애는 피로의 가장 큰 원인으로 작용하고 있는데, 수면시간은 집중력, 의사결정 능력 반응시간 등의 수행능력에 영향을 주고 이러한 수행능력이 떨어지는 상태를 피로 상태로 나타낼 수 있다[1, 31].

피로 수준 측정 방법에는 설문조사, 생화학/생리학적 검사, 반응속도 테스트 등이 있다. 설문에 의한 방법은 다원적 피로 척도 (Multidimensional Fatigue Inventory, MFI)[2]의 내용을 토대로 피로 측정 도구의 개발에 통계적으로 유의미한 결과를 보였다[3]. 이외의 피로 측정 방법으로는 Actigraphy를 활용한 일주기리듬 교란 현상에 대해 측정하여 평가하는 경우가 많다. 또 심박수, 맥파 등의 지표를 활용하여 측정하는 ECG(ElectrocardioGraphy) 방법, 피부의 온도와 같은 지표를 이용하는 생태순간평가 (Ecological Momentary Assessment) 방법이 있으며 혈액, 타액 등의 채취를 통해 측정하는 등의 생리학적 방법이 있다[4, 5, 32].

위와 같이 피로 측정에는 다양한 방법이 있지만, 주관적 판단이거나 개인의 상태와 상황에 따라 편차가 존재하여 피로도 산정에 절대적 기준으로 사용하기는 어렵다. 또한 조사에 기반한 피로 측정방법은 시간과 비용이 소요되고 식사 또는 운동, 감염 등 피로 측정에 영향을 미치는 요소들이 많다[6].

최근에는 피로와 관련된 데이터를 수집하고 딥러닝 모델을 학습하여 피로를 측정하는 연구가 진행되고 있다. 대표적으로 센서를 통해 신체의 움직임 데이터를 수집하여 피로를 측정하거나, 운전자를 지속적으로 촬영하여 눈의 깜빡임 등 동공, 홍채의 형태변화를 학습하여 피로를 측정하는 연구들이 있다[33-35].

본 연구는 주관적 피로도 데이터를 활용하여 딥러닝을 활용한 영상 분석의 피로 수준 측정의 가능성을 파악하고자 하였으며, 이후 정량적 데이터의 피로도 분류에 확장할 수 있는 모델을 제시하고 검증을 시도했다.

2. Video Analysis using Deep Learning Models

딥러닝을 기반으로한 영상 분석 기술들은 다양한 분야에서 연구되어 왔다. 일반적으로 영상은 여러 개의 프레임이 합쳐져 만들어진 데이터로 영상 분석 기술을 응용하여 분석된다. 영상 분석 시 사용되는 대표적인 딥러닝 모델은

Convolutional Neural Network (CNN)이며, 영상 분석은 시계열 정보를 포함하고 있어서 여기에 Recurrent Neural Network (RNN) 모델을 응용한다.

영상 분석에서 활용할 수 있는 CNN모델은 VGG[7], GoogleNet(Inception)[8], ResNet[9], Xception[10], MobileNet[11], DenseNet[12], EfficientNet[13], ViT[14], CoAtNet[15] 등 다양한 모델이 이상적인 모델로 발표되고 있으며, 최근 영상 분류(Classification) 문제에서 가장 성능이 좋은 ViT, CoAtNet모델은 자연어처리 분야에서 사용되는 Transformer[16]를 영상 분류 문제에 응용한 것이다. 영상 데이터는 일반적인 데이터와 다르게 공간적인 요소가 특징 중 하나이므로, CNN모델은 이를 포함하여 데이터의 특징을 추출하므로, 영상 분석에서 활용하기에 적합하다.

RNN모델은 LSTM[17], GRU[18]와 같은 은닉층을 활용하여 RNN 모델을 구성할 수 있다. RNN모델은 시퀀스(Sequence)가 입력, 출력으로 사용되는 방법으로 필기인식, 음성인식과 같은 시계열 정보를 포함하는 데이터에 주로 사용한다.

본 연구에서는 알려진 영상 분석 알고리즘을 피로에 적용할 수 있는 방법으로 조합하여 실험 및 검증을 통해 실제 활용할 수 있는 모델들을 중점으로 연구했다.

3. Finding Meanings on Videos

영상에서 특정 의미에 맞는 프레임을 찾아내기 위한 연구들은 영상기반의 행동 인식을 주로 다루고 있다.

초기에는 얼굴의 특징을 분류하는 연구가 주로 진행 되었는데 얼굴의 정면 이미지를 활용하는 기존 HOG 방법[19]으로 시작하여 자연스러운 영상을 이용한 분석을 추구하고 있다. UCF101 데이터셋[20]은 101개의 행동 클래스와 13,000개의 클립과 27시간의 동영상 데이터가 포함되어 있다. 이를 통해 사람의 행동을 인식하는 기술이 연구되고 있다. 대표적으로 Convolution 3D를 활용한 기법[21]이 있다.

본 연구에서는 영상에서 특정 의미에 해당하는 프레임을 찾아내는 것이 아니라, 영상 전체를 구성하는 프레임에서 특징들을 찾아내고 이를 활용하여 피로도를 측정하는 것에 목표를 두었다.

III. Frontal Face Video Analysis Models for Detecting Fatigue States

영상 데이터에 피로와 관련된 정보가 주어져 있다고 가정할 경우, 영상분석을 통한 피로 정보 적합은 분류모델이 된다. 일반적인 형태의 피로 분류 모델은 Fig. 1과 같다.

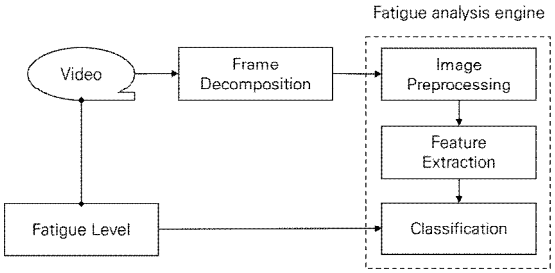


Fig. 1. Basic process of fatigue classification engine

이 모델은 먼저 영상을 화상으로 분해하는 과정을 거치게 되는데, 영상을 이루는 화상을 프레임(Frame)이라고 칭한다. 영상을 프레임 단위로 분해하는 과정에서 사용할 수 있는 방법은 무작위 추출, 시간 단위 무작위 추출과 Key frame 추출 방법 등이 있다.

이렇게 분리된 프레임들을 화상 분석 기법을 통해 분석하게 되는데, 일반적으로 특정 부분을 객체로 인식하는 등 전처리 과정을 거쳐 해당 화상에 존재하는 주요 부분 (얼굴의 경우 눈, 코, 입의 위치) 등 요소를 추출 (Feature extraction)하고, 이를 집합하거나 시계열적으로 종합하여 분석한 결과를 이용하여 최종적인 영상정보의 분류를 시행하게 된다.

1. Object-Feature Model

Object-Feature 모델은 전술한 영상 분석의 일반적인 형태를 따라 분석을 진행한다. Fig 2는 Object-Feature 모델의 구조를 나타낸다.

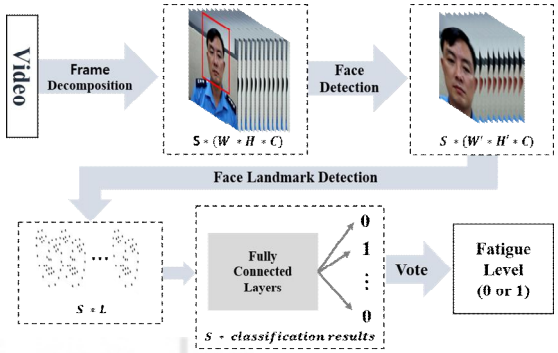


Fig. 2. Object-Feature model framework

Object-Feature 모델의 구체적 분석과정은 다음과 같다.

(1) 먼저 영상을 프레임 단위로 분해한다. 이때 하나의 영상은 다음과 같은 화상의 집합으로 분해된다.

$$M \approx \{S * (W * H * C)\}$$

S : 한 영상이 분해되는 프레임의 수, $0 < S \leqq s*f$
(s : 영상의 길이 (초), f : 영상의 초당 프레임 수)
W : 수평 해상도 (픽셀), $W > 0$
H : 수직 해상도 (픽셀), $H > 0$
C : 컬러 처리를 위한 채널 설정 (기본값: 3)

(2) 분해된 각 프레임에서 얼굴 부분의 영역을 검출한다. 얼굴 영역은 프레임의 해상도 내에서 추출되며, 주로 객체 인식 모델을 사용하게 되는데 대표적으로는 R-CNN[22], YOLO[23], SSD[24]등이 있다. 검출된 얼굴 영역은 다음과 같이 나타낼 수 있다.

$$M \approx \{S * (W' * H' * C)\}$$

W' : 검출된 얼굴의 수평 해상도 (픽셀), $0 < W' \leqq W$
H' : 검출된 얼굴의 수직 해상도 (픽셀), $0 < H' \leqq H$

(3) 검출된 얼굴 부분에서 분석에 사용될 주요 좌표 리스트를 추출한다. 본 연구에서는 얼굴 특징점 좌표 검출(Face landmark detection) 방법[25]을 이용하여 얼굴의 주요 좌표들을 추출하였으며, 다음과 같이 나타낼 수 있다. 이때 |L| = 68 로써, 하나의 얼굴 화상에서 68개의 얼굴 주요 특징점 좌표를 추출한다.

$$M \approx \{S * (L)\}$$

$L = (L_W, L_H)$: 얼굴의 주요부분 좌표, $|L|>0$

(4) 추출된 요소들을 딥러닝 알고리즘을 통해 주어진 피로 수준과 적합시키고, 이를 누적하여 최종적으로 피로 수준을 판단하는 분류기에 연결한다. 이 분석엔진의 구축 방법은 다양한데, 얼굴의 주요 좌표 리스트를 입력으로 사용하는 데이터의 특성상 완전연결층(Fully Connected Layer; FCL)을 이용한 분류기를 이용했다. 기본 구조는 Fig 3과 같다.

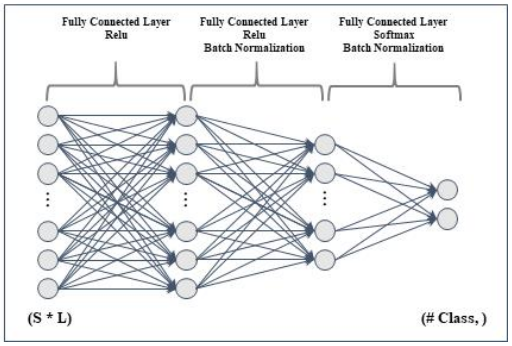


Fig. 3. Fully connected layer based classifier

학습 후 테스트 과정에서는 영상을 프레임으로 분해한 후 학습된 모델을 통해 각 프레임의 피로 수준별 확률을 도출하고 도출된 피로 수준 중 가장 많은 값이 최종 피로 수준으로 반환된다.

2. CNN Model

CNN 모델은 프레임 정보를 이용하여 피로 수준을 분류하고자 하는 모델로서, 객체 인식은 사용하지 않고 요소 추출은 별도의 방법이 아닌 모델 내에서 처리하도록 한 것이다. Fig. 4는 CNN 모델의 구성을 나타낸다.

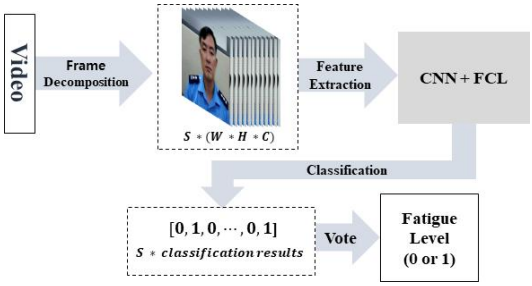


Fig. 4. CNN model framework

CNN 모델의 구체적 분석과정은 다음과 같다.

(1) 먼저 영상을 프레임 단위로 분해한다. 이는 모델 1과 동일하다.
(2) CNN을 이용하여 요소를 추출하고 이를 완전연결층으로 보내 주어진 피로 수준과 적합시키고, 이를 누적하여 최종적으로 피로 수준을 판단한다.

프레임의 요소를 추출하기 위해서는 CNN 계열 모델이 사용되며, 모델의 구조에 따라 그 요소는 벡터 형식으로 추출된다. 이 벡터는 완전연결층으로 이루어진 분류기로 바로 입력되어 주어진 피로 수준을 적합시키도록 하며, 기본 구조는 Fig 5와 같다.

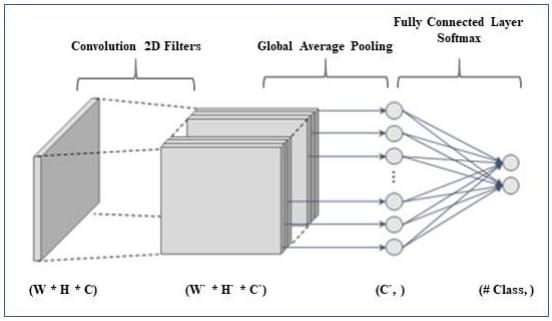


Fig. 5. CNN+FCL based classifier

CNN 모델에서 완전연결층을 구현하는 방식으로 Global Average Pooling (GAP)[30] 층을 사용하였는데, 이는 Object-Feature 모델과 같은 전통적인 완전연결층을 사용할 때 발생하는 폭발적인 파라미터 수의 증가로 인한 과적합을 예방한다. GAP 층은 추출된 특징 맵의 크기에 관계없이 각 컬러 채널들을 포함되는 값들의 평균으로 대체하므로 과적합 방지와 공간정보의 반영을 통해 분류 성능 향상을 기대할 수 있다.

학습 후 테스트 과정에서는 영상을 프레임으로 분해한 후 학습된 모델을 통해 각 프레임의 피로 수준별 확률을 도출하고 도출된 피로 수준 중 가장 많은 값이 최종 피로 수준으로 반환된다.

3. CNN-RNN Model

CNN-RNN 모델은 프레임 정보와 그 시간적 변화를 반영하고자 하는 모델로서, 객체 인식 미사용 및 요소 추출을 모델 내에서 하는 점은 모델 2와 같으나 추출된 요소들을 결합하여 분석하기 위해 RNN 기반 모델을 사용하는 것이 다르다. Fig. 6은 CNN-RNN 모델의 구조를 나타낸다.

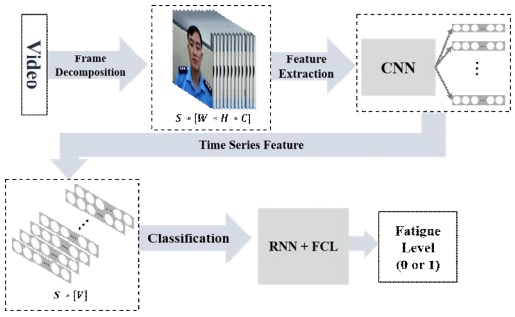


Fig. 6. CNN-RNN model framework

CNN-RNN 모델의 구체적 분석과정은 다음과 같다.

(1) 먼저 영상을 프레임 단위로 분해한다. 방식은 이전 모델과 유사하나, 분해된 프레임의 시간적인 전후관계가 포함되어야 하므로 프레임 분해 결과가 리스트로 반환된다.

$$M \approx [S * (W * H * C)]$$

(2) 분해된 프레임 리스트를 CNN 모델을 이용하여 요소로 추출한다. 이때 추출된 요소는 벡터(V) 형식을 띄게 된다.

(3) 프레임 벡터 리스트를 RNN 모델을 이용하여 시간적 변화와 함께 주어진 피로 수준과 적합시키고, 최종적으로 피로 수준을 판단한다. 이때 완전연결층에 적합시키기 위한 RNN 출력 사용치 (Sequence)는 Ts 만큼이며, RNN 모델의 기본 구조는 Fig 7과 같다.

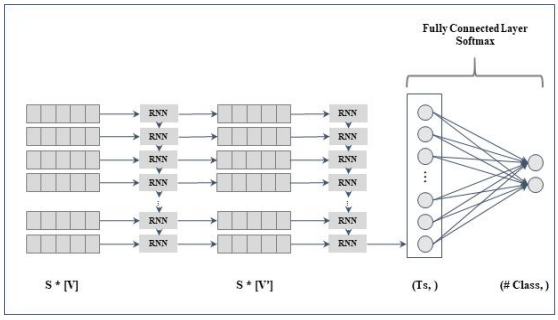


Fig. 7. RNN+FCL based classifier

CNN을 통해 추출된 프레임의 요소는 RNN 모델의 특정된 순서만큼 동시에 입력되며, 완전연결층으로 이루어진 분류기를 통해 피로 수준과 적합한다.

학습 후 테스트 과정에서는 다른 모델들과는 다르게 해당 영상을 학습된 모델을 통해 포함된 프레임 전체의 피로 수준 확률을 도출하게 된다.

IV. Experiments

1. Dataset

실험을 위해 사용한 데이터는 피로 관련 영상 및 생체 데이터의 수집 시스템 [26]을 통하여 수집된 데이터로서, 피험자의 영상 데이터와 피로도를 분석하여 도출된 피로 수준이 상정되어 있다. 각 데이터는 특정 시나리오와 대본대로 1분간 피실험자를 촬영한 영상으로서 대체로 피실험자의 정면에서 촬영되었으며, 피로도는 피실험자의 주관적인 판단하에 가장 좋음(1), 좋음(2), 보통(3), 나쁨(4), 가장 나쁨(5)의 피로 수준으로 조사되었다.

본 연구를 위하여 전체 데이터셋 중 피로 수준 1과 5에 해당하는 영상을 사용하여 실험했다. 총 533개 영상을

80%, 10%, 10%로 나누어 학습, 검증, 테스트 데이터셋으로 사용하였으며, 검증 데이터셋으로 모델별 하이퍼파라미터를 조정하여 가장 적합한 모델을 찾아내도록 하였고, 최종적으로 테스트 데이터셋을 이용하여 제안한 피로도 측정 모델들의 성능을 검증했다.

영상에서 프레임을 추출하는 방법으로 구조적 임의추출 방법 (Structured random sampling)을 이용하였는데, 한 영상의 전체 시간을 n초의 단위로 분할하고 각 단위에서 임의의 프레임을 추출했다. 즉 실험대상의 모든 영상은 n개의 프레임 집합으로 전처리되었다. (S = 60)

2. Algorithms for Experiments

Object-Feature 모델은 얼굴 영역의 검출을 위한 모델에서는 객체 인식 모델인 YOLOv3[27]를 사용하여 WIDER FACE 데이터셋[28]으로 얼굴을 학습시켜 얼굴 검출에 활용하였으며, 얼굴 특징점 추출에서는 Dlib 라이브러리의 Regression Tree Ensemble[29] 모델을 활용했다. 프레임에서 총 68개의 얼굴, 눈, 코, 입 좌표(x,y)를 추출하여 분류기의 입력으로 사용했다. 학습에 사용된 분류기는 Fig. 8와 같다.

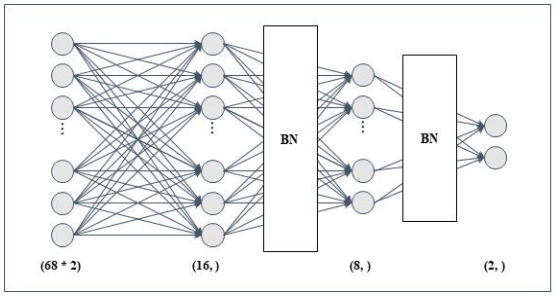


Fig. 8. FCL layer of Object-Feature model

68개의 좌표를 136개의 하나의 벡터를 입력으로 사용되고, 완전연결층 3개와 BN (Batch Normalization)층 2개로, 각 뉴런은 16, 8, 2개로 구성되어 마지막층에서 softmax 함수를 적용하여 주어진 하나의 피로수준으로 분류했다.

CNN 모델은 CNN에서 완전연결층이 결합된 형태로 Fig. 9과 같다.

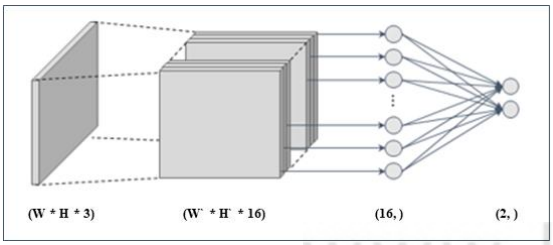


Fig. 9. CNN+FCL layers of CNN model

하나의 프레임(H, W, 3)크기를 입력으로 사용되며, Convolution 2D 층은 1개, 필터 개수는 16개, 커널크기는 (3, 3)으로 프레임의 요소를 추출하도록 하고, GAP(GlobalAveragePooling) 층으로 다차원 배열을 1차원으로 만든 후 완전연결층 1개로 피로도를 분류하도록 했다.

CNN+RNN 모델은 최대한 특징을 잘 추출할 수 있도록 현재 최적모델로 알려진 EfficientNetB7을 이용하여 프레임별로 2,560 크기의 1차원 배열을 도출하였고, 동영상 별로 추출된 프레임들을 하나의 시퀀스(Sequence)로 묶어 Fig. 10과 같이 RNN모델로 연결했다.

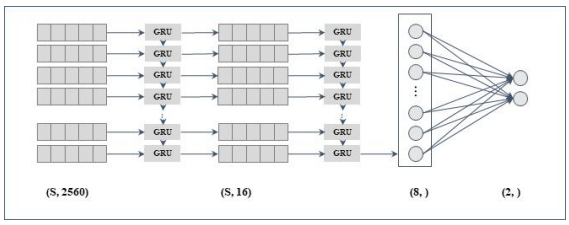


Fig. 10. RNN+FCL structure of CNN-RNN model

GRU층 2개로, 첫 번째 층에서 입력과 같은 시퀀스와 16개의 유닛을 출력(S, 16)하고, 두 번째 층에서는 시퀀스가 아닌 하나의 벡터로 8개의 유닛을 출력하며, 완전연결층 1개로 학습하여 피로도를 분류하도록 했다.

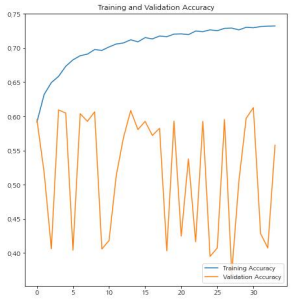
각 모델 학습 시 사용한 손실함수는 cross entropy를 사용하였으며, 방식은 식 (1)과 같다. 정답(y_i)과 예측값($\log(\hat{y_i})$)을 곱하여 손실함수를 최소화하는 방향으로 학습하도록 했다.

$$CE = -\sum_{i=1}^{N} y_i \log(\hat{y_i}) \qquad (1)$$

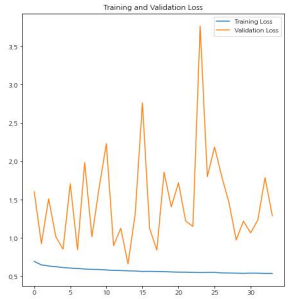
모델의 학습 및 테스트는 리눅스 기반의 GPU 탑재 단일 워크스테이션에서 Python 3.7 기반의 TensorFlow-Keras 환경에서 진행하였으며, 성능 평가의 척도로 실험대상 피로 수준별 정밀도 (Precision), 재현율 (Recall)과 피로 수준 전체에 대한 정확도 (Accuracy)를 사용했다.

3. Experimental Results and Discussions

모델별 학습 및 검증 단계에서의 정확도 및 손실 (loss)의 변화는 Fig 11과 같다.

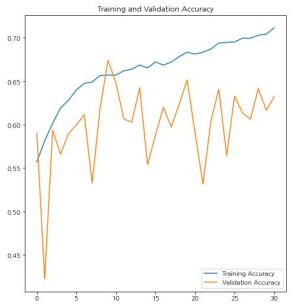


a-1. Accuracy

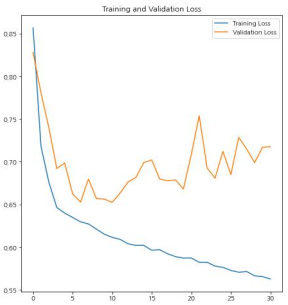


a-2. Loss

a. Object-Feature model

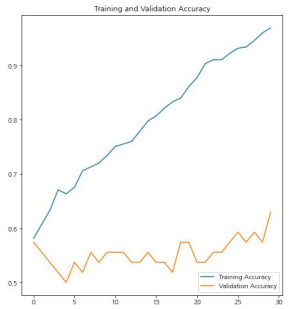


b-1. Accuracy

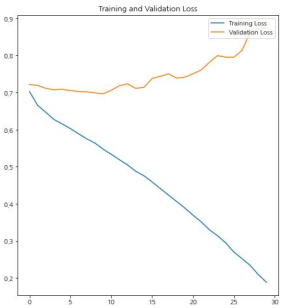


b-2. Loss

b. CNN model



c-1. Accuracy



c-2. Loss

c. CNN-RNN model

Fig. 11. Training / testing accuracies and losses

실험 결과 모델 모두가 30 에포크(epoch) 이전에 검증 정확도는 빠르게 수렴하고, 학습 손실은 지속적으로 하강하며 검증 손실은 상승하는 과적합의 양상을 보였다.

Object-Feature 모델은 학습 진행에 따라 검증 정확도는 60%를 상회하고 있으나 검증 손실이 변동성이 크고 손실값이 1~3.5로서, 학습된 모델이 검증용 데이터셋을 이용한 효과를 검증하는데 제한됨이 관찰되었다.

CNN 모델과 CNN-RNN 모델은 학습 진행에 따라 검증 손실이 줄어들다가 빠르게 상승하는 과적합 양상이 관찰되었다. 검증 손실은 CNN 모델이 0.65~0.85, CNN-RNN 모델이 0.7~0.9이며, 모두 Object-Feature 모델보다 비교적 낮게 산출되었다. CNN 모델은 CNN-RNN 모델에 비해 검증 정확도와 손실의 변동성이 크게 나타나 모델이 불안정하다고 볼 수도 있으나, CNN-RNN 모델은 학습 및 검증단계에서의 정확도와 손실의 괴리가 큰 것으로 관찰되어 입력 데이터의 수준에 비해 모델이 복잡하여 과적합이 심화되었다고 볼 수 있다. 따라서 실험 데이터셋에서의 성능은 검증 정확도의 최댓값이 크고 검증 손실의 최솟값이 작은 CNN 모델이 높은 분류성능을 보일 것으로 예측되었다. Fig 12는 실험을 통해 분류된 이미지의 예시이다.

a. Non-fatigue state

b. Fatigue state

Fig. 12. Example images

테스트 데이터를 통해 각 모델을 실험한 결과 도출된 성능은 Table 1에 나타냈다.

Table 1. Testing results

Model	Class	Precision	Recall	Accuracy
Object-Feature	1	0.6154	0.5000	0.5185
	5	0.4286	0.5455	
CNN	1	0.6591	0.9062	0.6667
	5	0.7000	0.3182	
CNN-RNN	1	0.6087	0.8750	0.5926
	5	0.5000	0.1818	

실험 결과 CNN 만을 활용한 모델이 가장 높은 성능을 나타냈다. 객체 인식과 요소 추출 모델의 경우 눈, 코, 입 등 얼굴의 주요 특징점만을 추출하여 피로도 분류에 사용하게되는데, 실험 결과로 볼 때 성능이 가장 낮게 나타났다. 이는 피로도의 경우 얼굴의 주요 특징점만이 아니라 얼굴의 형태, 몸의 형태, 피부의 색 등 전체적인 환경을 포함하여 산정됨을 시사한다고 하겠다.

CNN-RNN 모델은 실험 효율을 위하여 영상의 샘플 프레임을 사용하였으나, 영상의 전체 프레임을 분석에 사용하였을 경우에도 효율의 향상치는 매우 적었다. 이는 실험 데이터의 특성상 1분간의 짧은 영상만을 수집하였으므로 프레임의 시계열적 특성이나 그 변화가 두드러지지 않았다고 추정할 수 있으며, 입력 데이터의 규모가 모델의 복잡도에 적합하지 않은 측면도 있다. 향후 상대적으로 긴

시간동안 수집된 영상을 분석할 경우 CNN-RNN 모델이 보다 효과적으로 피로도를 추정할 수 있는 방법일 가능성이 있다. 또한 향후 낮은 Recall 값과 정확도를 개선하기 위해서는 각 클래스의 차이를 더욱 분명히 할 수 있는 데이터 수집 방법이나 영상 신호 전처리 과정을 적용해야 할 것이다.

V. Conclusions and Future Works

본 연구에서는 영상 데이터를 분석하여 피로 수준을 분류하는 딥러닝 기반의 모델을 제안했다. 특히 화상 분석에서 통상적으로 사용되는 객체 (얼굴) 인식, 요소 추출과 함께 영상 데이터의 시계열적 특성을 고려하여 방법론을 교차한 3가지 모델을 제시하였으며, 실험을 통하여 각 모델의 특성과 효과를 관찰했다.

3가지 모델 중 가장 성능이 좋은 CNN 모델의 경우 0.67의 정확도로 피로도를 분류하였다. 실험 결과 피로도 분류에 영상 분석 기반 모델이 적용 가능함을 확인할 수 있었으며, 모델별 학습 및 검증 절차 분석을 통해 영상 데이터 특성에 따른 모델 적용방안을 확인할 수 있었다.

본 연구에서 사용된 영상 데이터의 피로도 수준은 피실험자가 주관적으로 판단하여 입력한 것이므로, 향후 생리학적 분석을 통한 객관적 피로수준을 이용하여 영상 분석 모델의 효과를 보다 정밀하게 측정하여야 하겠으며, 키프레임 등 영상정보의 효과적인 전처리 방법 및 모델의 기능 확장 등 보다 효과적인 피로도 분류 엔진의 구성을 위한 연구가 필요하다.

ACKNOWLEDGEMENT

This research is supported by Civil-Military Dual Use Technology Development Work (No. 20-CM-BD-13) of Institute of Civil Military Technology Cooperation (ICMTC), funded by ROK Ministry of Trade, Industry and Energy and Defense Acquisition Program Administration.

REFERENCES

[1] Dinges, D. F. 1995. An overview of sleepiness and accidents. Journal of sleep research, 4: 4-14.

[2] Smets, E.; Garsssen, B; Bonke, B. d.; and De Haes, J. 1995. The Multidimensional Fatigue Inventory (MFI) Psychometric Qualifies of an Instrument to Assess Fatigue. Journal of Psychosomatic Research, 39(3):315-325.

[3] Lee, Y.; Shin, S.; Cho, T.; Yeom, H.; and Kim, D. 2021. An Experimental study on Self-rated Fatigue Assessment Tool for the Fatigue Risk Groups. In Proceedings of 2021 KMIST Conference, 1755-1756.

[4] Choe, J.-H,; Antoine, B. S. R.; and Kim, J.-H. 2014. Trend of Convergence Technology between Healthcare and the IoT. Information and Communications Magazine, 31(12):10-16.

[5] Kim, K.; and Lim, C. 2016. Wearable Health Device Technology in the IoT Era. The Korean Institute of Electrical Engineers, 65(11): 18-22.

[6] Kim, D. 2020. A Study on the Pilot Fatigue Measurement Methods for Fatigue Risk Management. The Korean journal of aerospace and environmental medicine, 30(2): 54-60.

[7] Simonyan, K.; and Zisserman, A. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.

[8] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; and Rabinovich, A. 2015. Going deeper with convolutions. In proceedings of the IEEE conference on computer vision and pattern recognition, 1-9.

[9] He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778.

[10] Chollet, F. 2017. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1251-1258.

[11] Huang, G.; Liu, Z.; van der Maaten, L.; and Weinberger, K. Q. 2017. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4700-4708.

[12] Howard, A. G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; and Adam, H. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861

[13] Tan, M.; and Le, Q. 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. International Conference on Machine Learning, 6106-6115. PMLR.

[14] Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.,; Heigold, G.; Gelly, S.; et al. 2020. An Image is Worth 16x16 Words:

Transformers for Image Recognition at Scale. arXiv preprint arXiv:2010.11929.

[15] Dai, Z.; Liu, H.; Le, Q.; and Tan, M. 2021. CoAtNet: Marrying Convolution and Attention for All Data Sizes. Advances in Neural Information Processing Systems, 34.

[16] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention Is All You Need. Advances in Neural Information Processing Systems, 30.

[17] Hochreiter, S.; and Schmidhuber, J. 1997. LONG SHORT-TERM MEMROY. Neural Computation, 9(8):1735-1780.

[18] Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. arXiv preprint arXiv:1406.1078.

[19] Dalal, N.; and Triggs, B. 2005. Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), volumne 1, 886-893. IEEE.

[20] Soomro, K.; Zamir, A. R.; and Shah, M. 2012. UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild. arXiv preprint arXiv:1212.0402.

[21] Kay, W.; Carreira, J.; Simonyan, K.; Zhang, B.; Hillier, C.; Vijayanarasimhan, S.; Viola, F.; Green, T.; Back, T.; Natsev, P.; et al. 2017. The Kinetics Human Action Video Dataset. arXiv preprint arXiv:1705.06950.

[22] Girshick, R.; Donahue, J.; Darrell, T.; and Malik, J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 580−587.

[23] Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. 2016. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779-788.

[24] Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; and Berg, A. C. 2016. SSD: Single Shot MultiBox Detector. In European conference on computer vision, 21-37 Springer.

[25] Sagonas, C.; Tzimiropoulos, G.; Zafeiriou, S.; and Pantic, M. 2013. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In Proceedings of the IEEE international conference on computer vision work-shops, 397-403.

[26] Yoo, S.; Kim, S.; Kim, D.; and Lee, Y. 2018. Development of Acquisition System for Biological Signals using Rasberry Pi. Journal of the Korea Institute of Information and Communication Engineering, 25(12): 1935-1941.

[27] Redmon, J.; and Farhadi, A. 2018. YOLOv3: An Incremental. arXiv preprint arXiv:1804.02767.

[28] Yang, S.; Luo, P.; Loy, C.-C.; and Tang, X. 2016. WIDER FACE: A Face Detection Benchmark. In Procedings of the IEEE conference on computer vision and pattern recognition, 5525-5533.

[29] Kazemi, V.; and Sullivan, J. 2014. One Millisecond Face Alignment with an Ensemble of Regression Trees. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1867-1874.

[30] Lin, M., Chen, Q., & Yan, S. (2013). Network in network. arXiv preprint arXiv:1312.4400.

[31] Lilleholt, L., Zettler, I., Betsch, C., & Böhm, R. (2020). Pandemic fatigue: Measurement, correlates, and consequences.

[32] Qiang, Y., Liu, J., & Du, E. (2017). Dynamic fatigue measurement of human erythrocytes using dielectrophoresis. Acta biomaterialia, 57, 352-362.

[33] Escobar-Linero, E., Domínguez-Morales, M., & Sevillano, J. L. (2022). Worker's physical fatigue classification using neural networks. Expert Systems with Applications, 198, 116784.

[34] Zhuang, Q., Kehua, Z., Wang, J., & Chen, Q. (2020). Driver fatigue detection method based on eye states with pupil and iris segmentation. Ieee Access, 8, 173440-173449.

[35] Yan, P., Sun, Y., Li, Z., Zou, J., & Hong, D. (2020). Driver fatigue detection system based on colored and infrared eye features fusion. Computers, Materials & Continua, 63(3), 1563-1574.

Authors

Simyeong Cha received the B.S. degrees in System Management and Engineering from Pukyong National University, Korea, in 2020. Cha is currently a Researcher at VAIV company. He is interested in ML/DL, AI, Computer Vision and Data Science.

Jongwoo Ha received B.S. degree in Industrial Engineering from Pukyong National University, Korea, in 2018, and M.S. degree in Management of Technology from Pukyong National University, Korea, in 2020, respectively. He completed Ph.D course in Industrial and Data Engineering from Pukyong National University, Korea, in 2022. Jongwoo Ha is currently a researcher of VAIV Company. He is interested in data science and Computer Vision, AI/ML.

Soungwoong Yoon received B.S. degree in Urban Engineering from Hanyang University, Korea, in 1996, and M.S. degree in Computer Science and Engineering from Korea National Defense University, Korea, in 2004, respectively. He completed Ph.D course in Informatics from Kyoto University, Japan, in 2014. Soungwoong Yoon is currently a principal researcher of VAIV Company. He is interested in analytics, AI/ML, and human intents.

Dr. Chang-Won Ahn is Director of Smart City Institute, VAIV Company. He is educated on Industrial Engineering, especially stochastic processes and queueing theory 1998 at KAIST (Korea Advanced Institute of Science and Technology), Korea. He has almost 20 years of experiences in system SW, BigData, AI technologies at ETRI. Since 2013, he has concentrated on the research area of Social Simulation or Computational Social Science. He is now charge of developing “Social Digital Twin” at VAIV Company, which is claimed as a foundation platform for dynamic & precision policy process for Smart City in the 4th Industrial Revolution era.