Term Project Report

recognition emotion

소프트웨어학과

2016312568

정 희 윤

1. Subject
   1. Explanation about Subject  
      영화에서는 종종 사람에게 연민을 느낄 정도로 정교한 감성지능 기술을 가지고 있는 인공지능이 등장하곤 한다. 이러한 영화 속 이야기는 더 이상 현실과 동 떨어진 주제가 아니다. 인간의 감성을 인지하고 해석할 수 있는 인공지능 기반의 시스템 개념을 감성 컴퓨팅 (Affective Computing: AC)이라고 한다. 이는 물리적 또는 감각적 자극으로부터 사람들 이 보이는 심리적인 반응을 인지하여 인간과 컴퓨터 간의 상호작용에 활용하기 위한 기술이다. 시장 조사 전문기관 가트너(Gartner)는 감성 컴퓨팅을 새로운 비즈니스 생태계 조성 가능성을 가진 잠재적 기술로서 인식했고, 2022년이면 인공지능이 인간의 감정을 이해해 지금보다 한층 더 적극적인 교감을 할 것이라고도 하였다.[[1]](#footnote-2) 이렇듯 감성 컴퓨팅 기술의 중요성이 점점 커지고 인공지능의 핵심기술로 여겨지고 있다. 감성 컴퓨팅 중 하나인 감정을 사람의 표정으로만 인식하는 것을 해보려고 한다. 사람의 표정으로만 감정을 높은 정확도로 맞추기는 힘들겠지만 목소리, 행동, 표정 등 여러가지를 결합하여 학습 시킨다면 높은 정확도로 맞출 것이라 예상된다.
   2. Why solve this Subject?  
      위와 같이 인공지능이 감정을 인식할 수 있다면 다음과 같은 장점들이 있다.   
      1) 인간의 감정을 이해하고 컨트롤하는 소프트웨어를 만들 수 있다.

2) 인간의 감정을 가진 휴머노이드 등의 로봇을 만들 수 있다.

3) 인간과 밀접하고 접근성이 좋다.

위와 같은 장점들 말고도 감정인식을 함으로써 얻는 이익은 훨씬 많을 것이라고

예상된다. 하지만, 흔히 CNN으로 진행하는 질병 분류 같은 경우에는 인풋이 사람에 따라

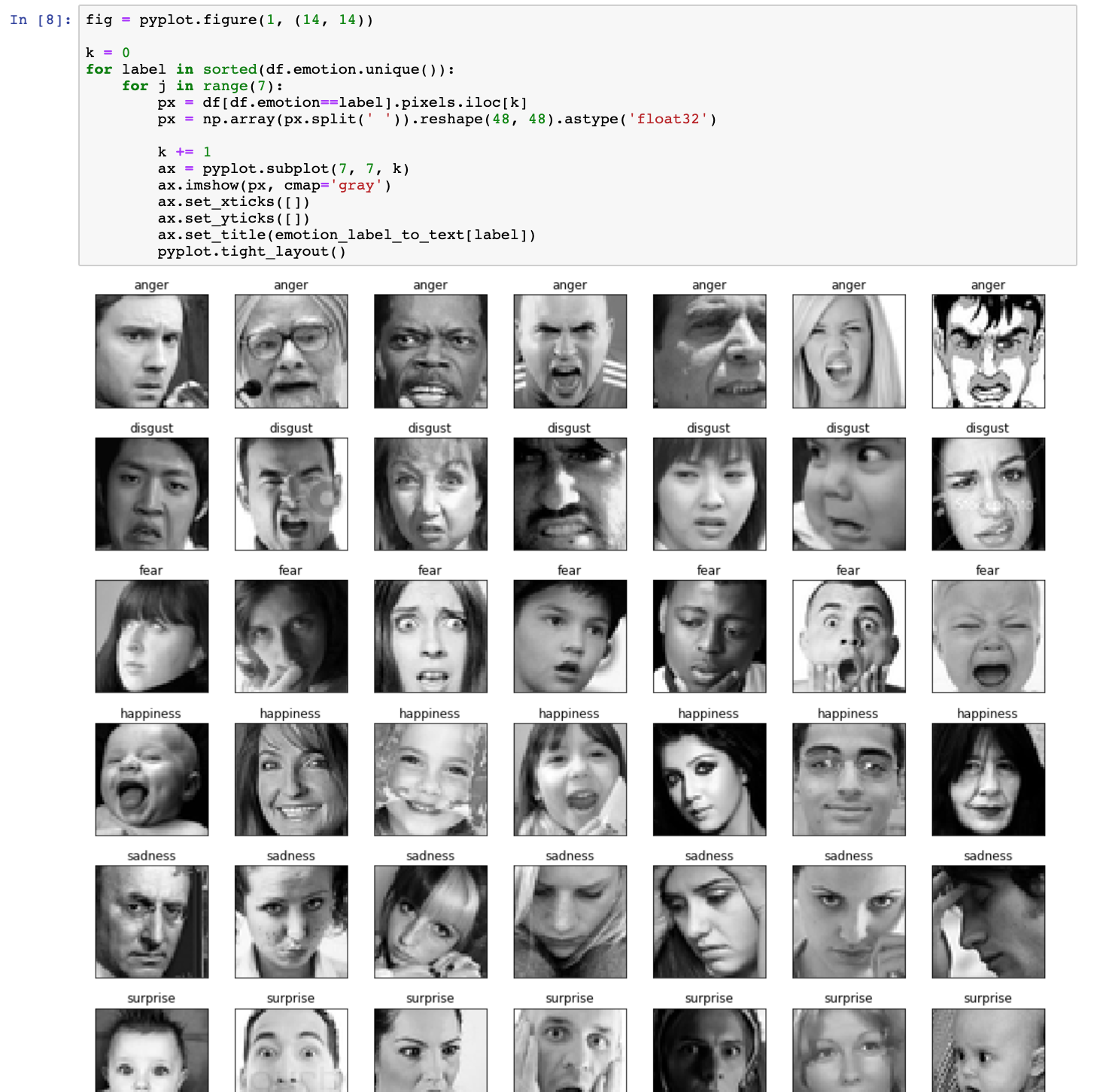
큰 차이가 없지만, 표정으로 감정 분류를 하는 것은 인풋이 사람마다 많이 다를 수

있기에 질병과 같이 큰 정확도를 기대하기는 어려울 것 같다. 하지만, 표정으로만

감정인식을 어느정도 할 수 있다면, 목소리 톤과 행동을 비롯한 여러가지 센서 정보들을

얻을 수 있다면 훨씬 정확한 감정인식 소프트웨어를 만들 수 있을 것이라고 생각한다.

그 시작점에 있는 ‘표정으로 감정 인식하기’를 주제로 정하였다.

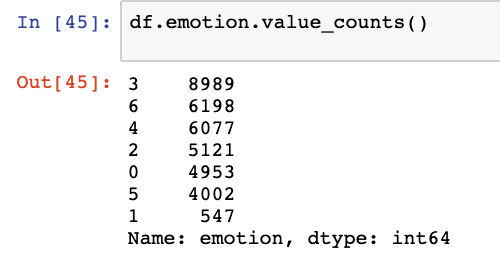
1. Dataset
   1. Explanation about Dataset  
      Dataset은 kaggle에서 진행한 fer2013 데이터를 사용하였다. 이 Dataset은 다음과 같이 흑백사진으로 이루어져 있다. 표정으로 감정인식을 할 때에 있어 조도와 배경에 따라 차이가 있을 수 있기에 고민하였으나 유명 기업들 ( ex) 마이크로소프트)에서 프로젝트를 진행 하였을 때에도 fer2013 데이터를 사용하였기에 이 Dataset을 채택하였다. 또한 Dataset이 36877개로 구할 수 있는 Dataset으로는 상당히 많은 편이기에 좋다고 생각하였다. 다음은 dataset의 일부분을 image화 시켰을 경우이다.
   2. Form of Dataset

Dataset은 감정 정보와 이미지를 픽셀단위로 바꾸어 csv에 저장되어 있는 다음과 같은 형태이다. Usage로 Traning과 Test가 있지만 이는 무시하도록 한다.  
(sklearn의 train\_test\_split을 사용하여 Train, Test Dataset을 정하도록 한다.)

* 1. Kinds of emotion

Dataset 에 있는 이미지들은 총 7가지의 감정으로 나누어져 있다. 각 감정들은 아래와 같이 화남, 역겨움, 공포, 행복, 슬픔, 놀람, 평범들로 나누어져 있다.

* 1. Counts of Dataset

Dataset 은 총 36877개 이며, 다음과 같은 형태로 저장되어 있다.  
각 감정에 대해서는 다음과 같은 개수로 존재한다. (각 감정 label에 대한 정보는 위의 사진과 같다.)  


* 1. Dataset used to training

Dataset에 따라 많은 학습을 시켜보았다. 어떠한 Dataset을 학습시킬지에 대해 종류별로 학습을 시도해보았는데 종류는 다음과 같다. Training Model은 아래 "3.2 Model”의 내용에 있는 Model을 사용하였다.

1. Dataset 전체
2. Dataset 중 감정에 따른 사진이 많은 상위3개 (Happiness, Neutral, Sadness)
3. Dataset 중 현저하게 적은 종류를 제외한 전체 (Disgust를 제외한 전체 Dataset)

위의 3개를 시도해 보았다.

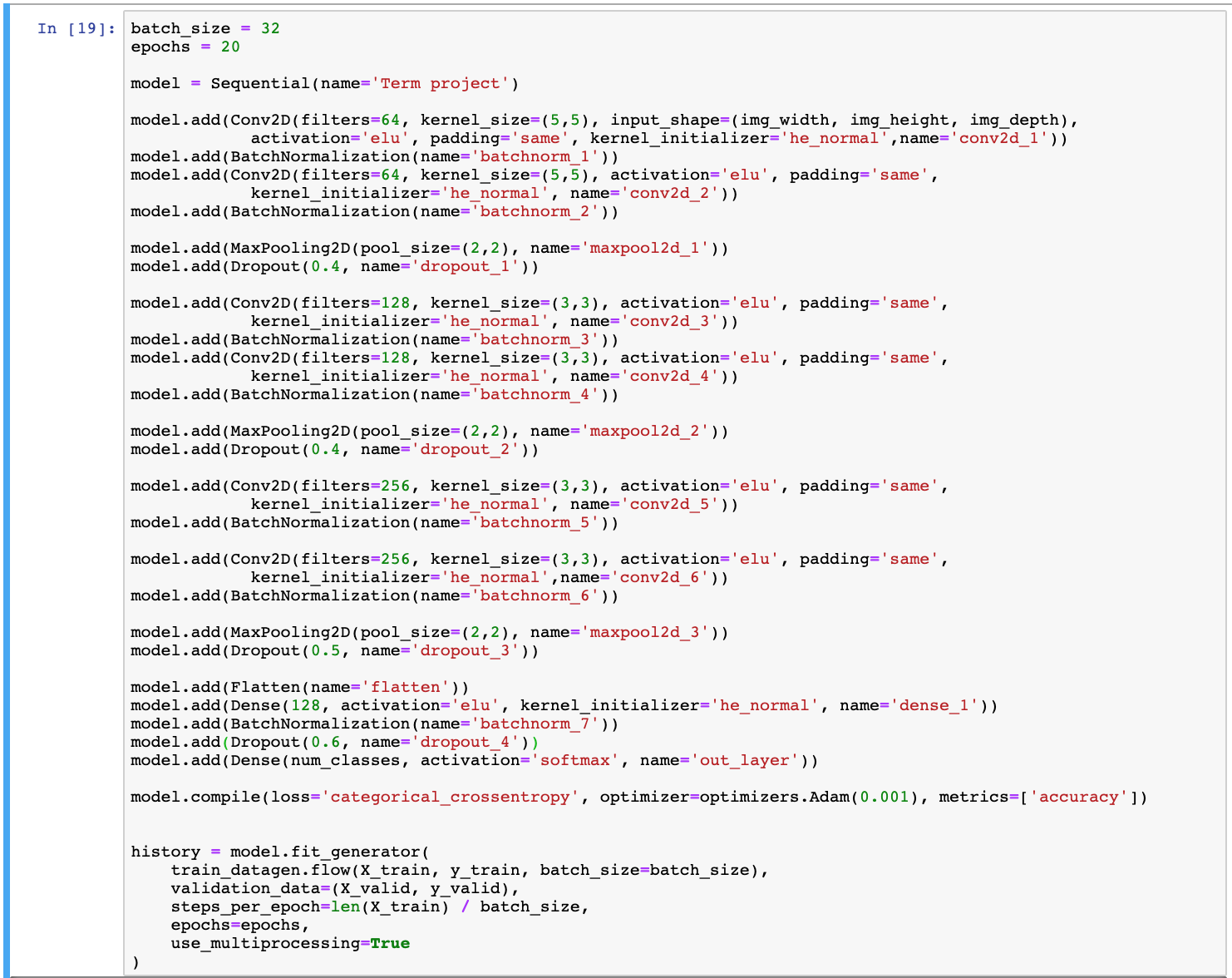
Dataset 전체로 학습을 시켰을 때에는 Disgust 데이터가 적어서 그런지 val\_acc가 50프로대로 상당히 낮으면서 학습시키는 시간이 매우 오래 걸렸다. Epoch를 100으로 설정하고 Early Stopping을 구현했음에도 Epoch가 80대로 매우 높고 80번 학습시키는데 50시간 이상으로 매우 오래 걸렸다. 따라서 전체 Dataset은 적합하지 않다고 생각한다.

Dataset 중 감정에 따른 사진이 많은 상위 3개로 학습을 시켰을 때에는 val\_acc도 80프로 대로 상당히 좋으면서 학습시키는 시간이 반 이하로 줄었다. 전체와 같이 Epoch를 100으로 설정하고 Early Stopping을 구현하였을 때 Epoch가 50대에서 끝났으며 학습시간은 30시간 이상으로 전체에 비해서는 상당히 좋았다. 하지만, 인간의 감정을 인식한다는 관점에서 Neutral, Happiness, Sadness 3가지로 분류해서 인식한다는 것은 처음에 말한 인간의 감정을 이해시킨다는 관점과 다른 것 같아 이 Dataset또한 적합하지 않다고 생각한다.  
 위와 같은 이유로 최종적으로 사용하는 Dataset은 ‘Dataset 중 현저하게 적은 종류를 제외한 전체 (Disgust를 제외한 전체 Dataset)’로 설정하였다.

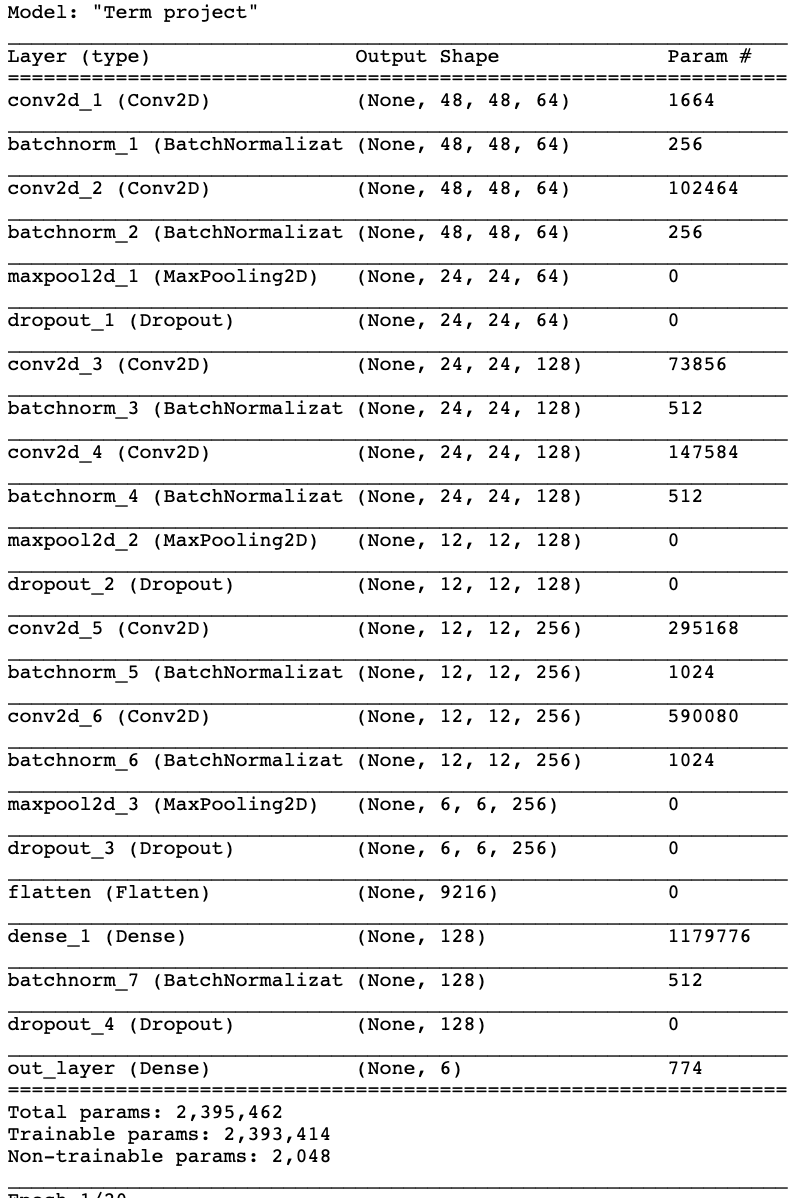
1. Algorithm
   1. Kind of Algorithm

감정을 인식하기 위해서는 여러 사진들을 감정에 따라 classification 해야 한다. Images를 classification하는데에 있어 CNN이 적합하다고 생각하여 CNN으로 진행하였다.

* 1. Model

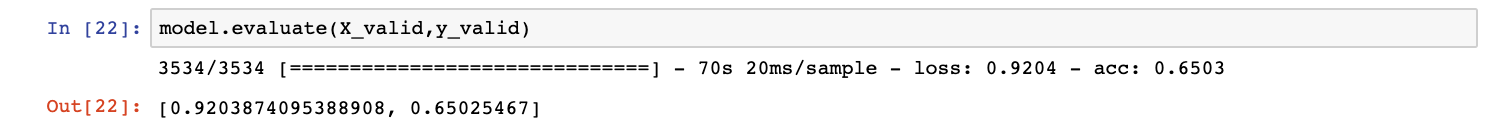
학습 모델은 다음과 같다.

위의 모델로 되기까지 여러 과정이 있었다. 처음에는 여러 기본 CNN모델 ( ex) AlexNet, VGG19 등)으로 테스트 해보았는데 정확도가 50대 중반정도로 상당히 낮게 나왔다. 정확도를 높이기 위해 여러 모델을 찾아보다가 최종적으로 취합한 모습이다. 또한 batch\_size를 64, 32로 해보았을 때 32가 경험적으로 정확도가 더 높았다. Activation은 relu와 élu Optimizer은 Adam과 NAdam으로 하여 테스트를 해보았는데 Activation이 elu일때 더 정확도가 좋았으며 Optimizer은 Adam과 NAdam 둘 다 차이가 거의 없었다. elu를 사용하기위해 initializer은 ‘he\_normal’을 사용하였다.

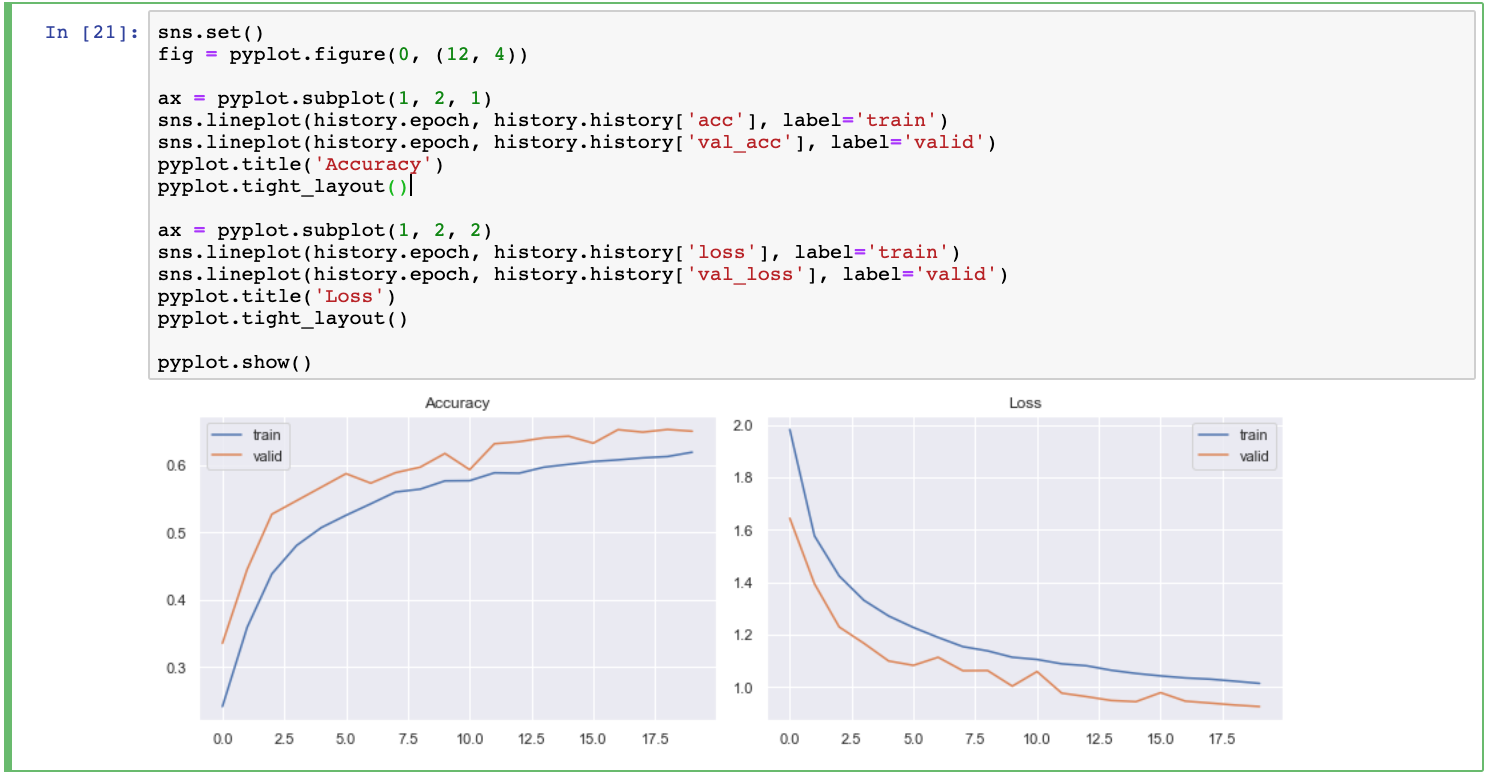
elu는 relu와 특성은 동일하나 dying relu문제를 해결해주어 테스트를 좀 더 잘 시킬 수 있지만 elu를 사용하면 relu exp()를 계산해주어야 하는 비용이 든다는 단점이 있다. 이 때문인지 학습을 시킬 때에 있어 val\_acc와 test 정확도는 올라갔으나 epoch 한번당 30분 이상씩 들어 시간 관계상 최종 진행할 때는 epoch를 20으로 하였다.

위는 Model의 summary()의 출력 값으로, 총 변수는 2395462이다.

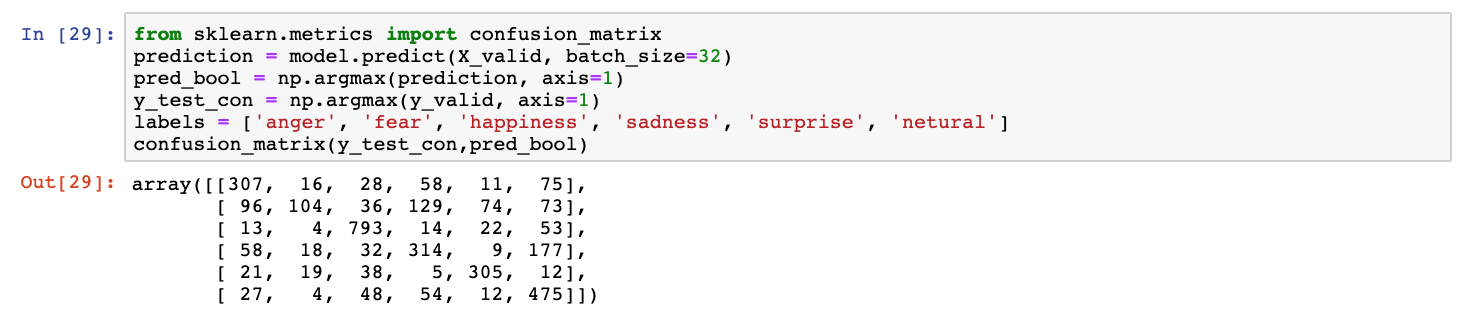
1. Result
   1. Result of test dataset

다음은 test dataset으로 평가하였을 때 결과이다.

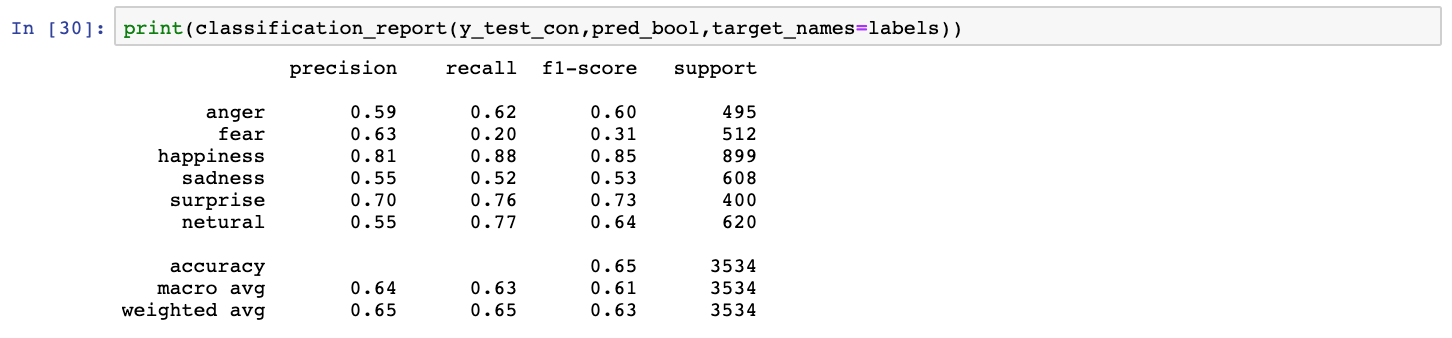
* 1. Graph of training history

다음은 학습시킬 때 history에 대한 것이다. 다음은 Epoch에 따른 model의 Accuracy와 Loss값을 나타낸 그래프이다.

* 1. Confusion matrix

다음은 Model의 Confusion\_matrix이다.

* 1. Classification report

다음은 model의 classification report이다.

* 1. Prediction

다음은 감정 별 사진과 실제 감정, 예측 감정 값을 나타낸 것이다.

* + 1.  Anger and fear
    2.  Happiness and sadness
    3.  Surprise and Natural

1. Result Analysis  
   ‘4. Result’ 항목에서 나온 결과들로 다음과 같은 내용을 알 수 있다.
2. 학습 결과로 약 65%의 accuracy를 얻을 수 있었다. 현재는 데이터 개수가 약 35000개 정도로 학습을 시켰는데 좀 더 충분한 dataset 으로 학습을 시킨다면 더 높은 정확도를 기대할 수 있을 것 같다.
3. History graph를 보면 Epoch가 증가함에 따라 loss는 점점 감소하고 Accuracy는 점점 증가하는 모습을 볼 수 있다. Train data는 꾸준히 loss가 감소하고 Accuracy는 증가하지만 Valid data는 loss가 감소하고 accuracy가 증가하는 추세이긴 하지만 약간의 편차가 있었다. 그래프의 기울기 증감을 확인해보면 Epoch를 더 높게 잡았다 해도 그리 많은 정확도 차이가 날 것 같지는 않다. Epoch를 100가까이한다면 60대 후반의 정확도가 나올 것이라 예상한다.
4. Classification Report의 f1-score를 보면 fear가 다른 부분들과 비교했을 때 현저히 낮고, sadness와 anger은 0.5~0.6으로 좋지 않고 나머지 부분들은 양호하게 학습이 이루어졌다고 볼 수 있다.
5. ‘4.5 Prediction’의 결과는 Test dataset에서의 prediction 일부분이다. Classification Report에서와 같이 감정상태가 fear인 경우는 잘 맞지 않은 것을 볼 수 있고, anger 또한 잘 맞지 않음을 볼 수 있다. 하지만 나머지 감정상태는 거의 다 맞추는 것을 볼 수 있다. 이에 대한 이유를 고민해보면, sadness case의 첫번째 사진이 잘못된 input이 들어감을 볼 수 있다. 이러한 잘못된 인풋이 학습시키는데 있어 잘못된 영향을 줄 수 있다. 또한 감정이라는 주제가 인간마다 판단하는 것이 다르고 명확히 감정을 나눌 수 없기에 정확도가 그리 높지 못한 것 같다.
6. Conclusion  
    Term project를 시작하고 ‘인간의 감정을 표정으로만 판단하는 프로젝트를 해보자’라고 주제를 정한 뒤 여러 시도가 있었다. 처음은 50%대 초반의 정확도에서 여러 모델구조와 Activation, Optimizer을 바꿔보며 65%로 정확도를 올렸다. 데이터셋이 있기에 학습시키고 결과를 도출하는데 금방 할 것이라는 생각과는 달리 CNN구조로 최대한 높은 정확도를 구하기위해 모델을 바꾸고 학습시켜보고 하는데 시간이 엄청 오래 걸렸다. 초반에는 학습시키는데 2일이 걸렸으며 마지막에는 Epoch가 20임에도 불구하고 학습시키는데 14시간이 넘게 걸렸다. 물론 장비가 좋지 않기에 이정도로 걸린 것이지만 여러 제약사항이 있어 많이 아쉬웠다. 결과를 도출할 때 ‘4.5 Prediction’ 내용을 도출하니 65%의 정확도가 낮지 않은 것이라는 것을 알 수 있었다. 인간인 내가 보았을 경우에도 명확히 감정이 드러나는 사진들도 있지만 아닌 것도 있었다. 오히려 내가 생각했을 때도 실제 데이터 값보다 prediction한 값이 더 맞는 것 같은 데이터들도 있었다. 인간이 이 dataset으로 감정을 판단한다면 80%정도라고 생각한다. 그렇기에 65%라는 결과는 절대 낮은 것이 아니라고 생각한다. 데이터 셋이 훨씬 풍부하다면 훨씬 나은 결과를 얻을 수 있을 것이라 생각한다. 또한 처음에 이 주제로 시작한 의도는 표정으로만 감정을 판단하는 것이 어느정도 나온다면 목소리, 행동, 표정 등으로 종합적인 판단을 한다면 매우 정확 해질 수 있다고 생각하기에 시작한 것이다. 65%라는 정확도를 얻었기에 목소리, 표정, 행동 등 여러 사항들을 종합해 본다면 매우 좋은 정확도를 얻을 수 있을 것이라 생각한다.

1. 넥서스커뮤니티, “우리는 인공지능과 사랑의 교감을 나눌 수 있을까?”, NEXUS Business & Culture Magazine-BUZZ, Vol.28, 2018. [↑](#footnote-ref-2)