다중 클래스 이미지 표정 분류

오명호 · 민송하 · 김종민*

동신대학교

Multiclass image expression classification

myung-ho Oh · Min-song-ha · Jong-min Kim*

Dongshin University

E-mail : dhajd444@gmail.com / thdgk1052@naver.com / dyuo1004@dsu.ac.kr

요 약

본 논문에서는 지도 학습에 기반을 둔 다중 클래스 이미지 장면 분류 방법을 제시한다. 데이터 세트에서 콘볼루션 뉴런 네트워크 모델에 학습시켜 다중 클래스 사람의 표정 장면 이미지를 분류하였으며, 실험에서는 최적화된 CNN 모델을 Google image 데이터 세트에 분류하여 유의미한 결과를 얻을 수 있었다.

ABSTRACT

In this paper, we present a multi-class image scene classification method based on map learning. We were able to learn from the convolutional neural network model in the dataset, classify facial scene images of multiclass people, and classify the optimized CNN model into the Google image dataset in the experiment with significant results.

키워드

multi class image sence classification method, transformation learning, image datasets

Ⅰ. 서 론

최근 컴퓨터 비전 기술의 발전에 따라 이미지 그래픽 처리 알고리즘 기술에 대한 연구가 급속하게 촉진되었으며 그중 이미지 분류에 대한 알고리즘이 영상 처리 기법에서 큰 관심을 받고 있다. 또한 이미지 처리 알고리즘은 객체 감지, 의미론적 분할법 등에 널리 사용되고 있으며 CNN 모델은 데이터 세트의 수와 레이어 층의 깊이에 따라 정확도는 높아지지만 시간이 더 많이 걸리고 컴퓨터 컴퓨팅 성능이 더 많이 필요하다. 또한 본 연구에서 사용된 합성곱 신경망은 다른 딥러닝 모델과 마찬가지로 다수의 은닉층을 포함하고 있기 때문에 모델 복잡도가 높고, 일정수준 이상의 성능을 가지는 모델을 구현하기 위해 충분한 양의 학습 데이터를 보유해야 할 필요가 있다. 예를 들어, 우수한 성능으로 잘 알려져 있는 합성곱 신경망 모델 중 하나인 VGGNet[1]의 경우에는 약 1억 3천 8백만개의 파라미터를 포함하고 있기 때문에, 학습을 수행하기 위해서는 많은 수의 학습 데이터가 필수적이다. 현실적으로 우수한 성능을 가지는 합성곱 신경망 모델을 구현할 수 있을 만큼의 학습 데이터를 실제 환경에서 확보하는 데 있어서 많은 제약조건이 존재하기 때문에, 이를 해결하기 위해 보유하고 있는 학습 데이터를 확대, 축소, 회전 시키는 등의 선형 변환 방식을 이용하여 데이터세트의 수를 늘려 해결한 뒤 연구를 진행하였다.

Ⅱ. CNN Model

연구에 사용된 CNN(Convolution Leural Network) 모델은 Convolution Layer와 Max Pooling 레이어를 반복적으로 stack을 쌓는 특징 추출(Feature Extraction) 부분과 Fully Connected Layer를 구성하고 마지막 출력층에 Softmax를 적용한 분류 부분으로 나뉩니다. 그림 1은 전형적인 CNN의 구조이다.

* corresponding author

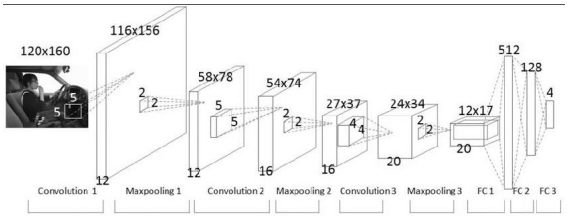
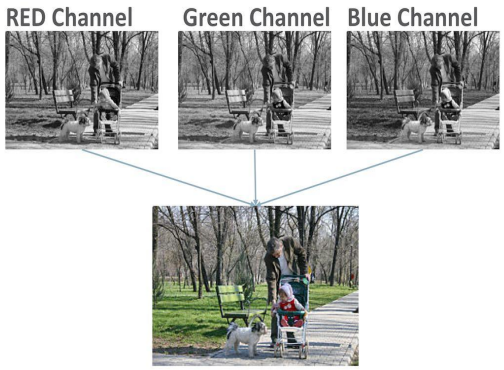


그림 1. CNN의 기본 구조

이미지 데이터를 입력받기 위해서는 이미지를 픽셀 단위로 수치화 시켜주어야 하며, 이미지 픽셀 하나하나는 실수의 값을 가진다. 그림 2의 컬러 사진은 천연색을 표현하기 위해서, 각 픽셀을 RGB 3개의 실수로 표현한 3차원 데이터이며, 컬러 이미지는 3개의 채널로 구성된다.



이미지 출처: https://en.wikipedia.org/wiki/Channel_(digital_image)

그림 2. 이미지 데이터의 채널

Convolution Layer는 합성곱 레이어 라고도 하며 이는 입력 이미지를 특정 Filter(Kernel)를 이용하여 탐색하면서 이미지의 특징들을 추출하고, 추출한 특징들을 Feature Map으로 생성한다. 피처맵은 필터가 전체 이미지를 Stride하며 모든 픽셀과 연산을 거쳐 나온 결과값이며,stride는 보폭이란 의미로 필터를 적용하는 간격을 정하게 된다. 필터의 크기와 같다.

Pooling 레이어는 컨볼루션 레이어의 출력 데이터를 입력으로 받아서 출력 데이터(Activation Map)의 크기를 줄이거나 특정 데이터를 강조하는 용도로 사용된다. 이렇게 만들어진 신경망 모델을 기반으로 머신러닝을 통해 새로운 이미지에 대한 데이터 정보를 분류해주게 된다. 또한 수식(1)을 통해 활성화 함수로 사용된 soft max는 입력받은 값을 출력으로 0~1사이의 값으로 모두 정규화하며 출력 값들의 총합은 항상 1이 된다. 분류하고 싶은 클래스의 수 만큼 출력으로 구성하고, 가장 큰 출력 값을 부여받은 클래스가 확률이 가장 높은 것으로 이용된다.

$$f(x)_i = \frac{e^{\vec{\xi}}}{\sum_{k=1}^{K} e^{x_k}} \text{ for } i = 1,...,K \qquad (1)$$

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 64, 64, 32)        896

max_pooling2d (MaxPooling2D  (None, 32, 32, 32)        0
)

dropout (Dropout)            (None, 32, 32, 32)        0

conv2d_1 (Conv2D)            (None, 32, 32, 64)        18496

max_pooling2d_1 (MaxPooling  (None, 16, 16, 64)        0
2D)

dropout_1 (Dropout)          (None, 16, 16, 64)        0

flatten (Flatten)            (None, 16384)             0

dense (Dense)                (None, 256)               4194560

dropout_2 (Dropout)          (None, 256)               0

dense_1 (Dense)              (None, 3)                 771

=================================================================
Total params: 4,214,723
Trainable params: 4,214,723
Non-trainable params: 0
```

그림 3. 신경망 구조

그림 3은 실험에 사용된 실제 신경망의 소스 코드이며, 컨볼루션레이어와 맥스풀링 레이어의 반복 구성으로 돼어 있는 구조이다.

III. 실험 및 결과

비교 실험을 사용하여 사람의 표정 이미지 분류 데이터 세트에 대해 실험을 수행했으며, 먼저 CNN 모델에 대한 훈련과 테스트를 진행했다. 실험결과 training sample은 batch size가 32, training step size가 batch당 32, epoch가 50의 값을 얻을 수 있었으며, early_stopping 함수를 통해 테스트 손실이 더 이상 떨어지지 않으면 훈련이 종료된다. 테스트에 사용되는 이미지는 이미지 URL을 입력하는 방식으로 분류 실험을 시킬 수 있으며, 최종적으로 사람의 표정에 대한 클래스를 분류할 수 있다.

그림 4. 이미지 분류 결과 출력화면

Ⅳ. 결론 및 향후 연구

본 논문에서는 이미지 학습을 기반으로 다중 클래스 이미지 분류를 할 수 있는 연구를 수행하였으며, 각각의 표정 이미지 데이터를 분류하기 위해 CNN모델을 사용하여 사람의 표정 이미지 데이터 세트를 분류하는 방법에 대해 제시하였다. 또한 기존 연구에 사용된 선형 변환은 정확도 면에도 완성도가 떨어지는 부분이 있으므로 다른 방법을 추가할 필요성이 있으며, 전이 학습에 대한 알고리즘을 추가시킨다면 고무적인 결과를 도출할 수 있을 것이다.

Acknowledgement

본 과제(결과물)는 2021년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다.(NRF-13453 41782)

References

[1] S. -Y. Shin, H. C. Lee, K. -S. Shin, H. -J. Kim, and J. -W. Lee “Image Scean Classification of Multiclass,” *Journal of Digital Convergence*, Vol. 17. No. 8, pp. 105-113, 2019.

[2] IHansoo Lee, J. Kim, J. Yu, Y. Jeong, and S. Kim, “Multi-class Classification using Transfer Learning based Convolutional Neural Network,” Vol. 29, No. 5, Oct. 2019

[3] E. W. Na and O. S. Kang, “A Method for Generating Metamorphic Test Cases for CNN Image Classification Models,”