

https://doi.org/10.7236/JIIBC.2020.20.1.225
JIIBC 2020-1-32

이미지 인식을 개선을 위한 CNN 기반 이미지 회전 보정 알고리즘

CNN-based Image Rotation Correction Algorithm to Improve Image Recognition Rate

이동규*, 선영규*, 김수현*, 심이삭*, 이계산**, 송명남***, 김진영****

Donggu Lee*, Young-Ghyu Sun*, Soo-Hyun Kim*, Issac Sim*,
Kye-San Lee**, Myoung-Nam Song***, Jin-Young Kim****

요약 이미지 인식 및 영상처리, 컴퓨터 비전 등의 분야에서 합성곱 인공신경망 (Convolutional Neural Network, CNN)은 다양하게 응용되고 탁월한 성능을 내고 있다. 본 논문에서는 CNN을 활용한 이미지 인식 시스템에서 인식을 저하시키는 요인 중 하나인 이미지의 회전에 대한 해결책으로써 CNN 기반 이미지 회전 보정 알고리즘을 제안한다. 본 논문에서는 Leeds Sports Pose 데이터셋을 활용하여 이미지를 임의의 각도만큼 회전시킨 학습데이터로 인공지능 모델을 학습시켜 출력으로 회전된 각도를 추정하도록 실험을 진행하였다. 학습된 인공지능 모델을 100장의 테스트 데이터 이미지로 실험하여 mean absolute error (MAE) 성능지표를 기준으로 4.5951의 값을 얻었다.

Abstract Recently, convolutional neural network (CNN) have been showed outstanding performance in the field of image recognition, image processing and computer vision, etc. In this paper, we propose a CNN-based image rotation correction algorithm as a solution to image rotation problem, which is one of the factors that reduce the recognition rate in image recognition system using CNN. In this paper, we trained our deep learning model with Leeds Sports Pose dataset to extract the information of the rotated angle, which is randomly set in specific range. The trained model is evaluated with mean absolute error (MAE) value over 100 test data images, and it is obtained 4.5951.

Key Words : Rotation Estimation, Deep Learning, CNN, Image Processing

1. 서 론

합성곱 인공 신경망 (Convolutional Neural Network, CNN)은 격자 구조 형태의 데이터 처리에 탁월한 성능을

보인다^{1), 2)}. 이점에 착안하여 2차원 이미지처리 및 실시간 영상처리 분야에서 활발한 연구가 진행되고 있다³⁾⁻⁵⁾. Deepmind의 연구에서는 인공지능 모델의 입력으로 84*84 픽셀의 이미지를 사용하고, 3개의 합성곱 계층과

*준회원, 광운대학교 전자융합공학과

**정회원, 경희대학교 전자공학과

***준회원, ㈜현대엔터테인먼트

****정회원, 광운대학교 전자융합공학과

접수일자: 2019년 11월 6일, 수정완료: 2020년 1월 6일

게재확정일자: 2020년 2월 7일

Received: 6 November, 2019 / Revised: 6 January, 2020 /

Accepted: 7 February, 2020

****Corresponding Author: jinyoung@kw.ac.kr

Dept. of Electronic Convergence Engineering, Kwangwoon Univ., Seoul, Korea

한 개의 퍼셉트론 계층을 사용하여 Atari 사의 비디오 게임의 플레이를 학습하는 영상인식 모델을 제안하였다^[3]. 최적화 방법으로는 인공지능 기법 중 하나인 강화학습을 기본적인 최적화 방식으로 사용하였다. 인공지능 모델이 영상인식을 통해 게임 환경을 파악하고 최적의 플레이를 선택하게 하여 높은 수준의 성능을 보였다.

또한, 12,000 개의 고해상도 이미지를 1,000 개의 분류 기준으로 분류하는 CNN 모델 구조를 제안하는 연구도 진행되었다^[4]. 해당 연구에서는 60 만 개의 파라미터를 갖는 CNN 구조를 학습시켜 오차율 15.3%의 성능을 보였다. 이는 ImageNet LSVRC-2010 대회에서 가장 우수한 성능으로, 방대한 양의 파라미터와 깊은 네트워크 구조를 통해 정밀한 분류모델을 구현하였다.

사람의 이미지를 학습하여 주요 관절 및 신체구조를 파악하고 이를 통해 운동 종류를 분류하는 CNN 모델을 설계한 연구가 있다^[5]. CNN 모델은 운동하는 사람의 이미지를 입력으로 받아 14개의 주요 관절 및 신체구조의 좌표를 학습하게 된다. 해당 연구에서는 3 개의 합성곱 계층과 2개의 퍼셉트론 계층으로 이루어진 네트워크를 제안하였다.

일반적인 인공지능 기반 영상인식 시스템은 이미지의 픽셀값을 추출해 이미지 처리를 진행한다. 이때 입력 이미지에서의 주요 인식 객체가 평행이동하거나 회전된다면, 인공지능 모델은 이를 새로운 이미지로 판단하게 되어 인공지능 모델의 인식을 현저히 떨어뜨리는 문제점이 있다. 이에 본 논문에서는 인공지능을 활용한 영상인식 시스템에서 인식을 개선하기 위해 이미지의 기울임을 보정하는 CNN 기반 regression 기법을 제안한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 앞서 언급한 다층 퍼셉트론, CNN에 대해 기술한다. 3장에서는 사용한 데이터와 실험 환경 및 파라미터들에 대해 설명하고 실험에 따른 결과를 제시한다. 마지막 4장에서는 결론을 제시하고 후속 연구에 대해 논의하며 마무리 한다.

II. 딥 러닝 알고리즘

딥 러닝은 머신러닝의 한 분야이며, 인간 두뇌의 학습 구조를 파악하고 탐구하는 분야이다. 이를 증가적으로 구현한 인공지능망은 입력층, 은닉층 그리고 출력층으로 이루어지며 계층적인 구조를 띤다. 특히, 은닉층의 깊이가 2층 이상인 인공지능망을 딥 러닝 네트워크라고 한다^[6].

1. 다층 퍼셉트론

다층 퍼셉트론을 이루는 단위인 퍼셉트론은 인공지능망을 이루는 기본 단위로써 인간의 뉴런을 증가적으로 구현한 것이다. 이때 퍼셉트론의 출력을 수식적으로 표현하면 다음과 같다.

$$y = f(W_j x_j + b_j). \quad (1)$$

위의 수식 (1)에서 y 는 퍼셉트론의 출력, f 는 퍼셉트론의 활성화함수, W_j 는 퍼셉트론의 입력 x_j 와 퍼셉트론의 연결계수, b_j 는 퍼셉트론에 더해지는 바이어스값을 나타낸다. 이러한 퍼셉트론을 다층구조로 구성한 것을 다층 퍼셉트론 이라고 한다^[6].

2. CNN

CNN은 인공지능망을 이루는 은닉계층으로 합성곱 계층을 사용한 인공지능망을 의미한다. 이때 합성곱 계층의 출력은 다음의 수식으로 나타난다^[1]. 이때 I 는 입력, K 는 합성곱 커널 또는 필터를 의미하며,

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n) K(m, n). \quad (2)$$

m 과 n 은 입력의 2차원 크기를 나타낸다. 그림 1은 수식 (2)의 과정을 보여주는 그림으로 CNN의 기본 연산 과정을 보여준다. 그림 1의 입력은 3행 3열 크기를 갖는 행렬이지만, 2행 2열 커널과의 합성곱 연산을 통해 2행 2열 출력으로 나오게 된다. 커널의 크기와 합성곱 연산 횟수에 따라 출력의 크기는 입력과 다르게 변하게 된다. 이를 방지하기 위해 입력 데이터 행렬에 0의 값을 이어 붙이는 풀링 기법을 이용하기도 한다.

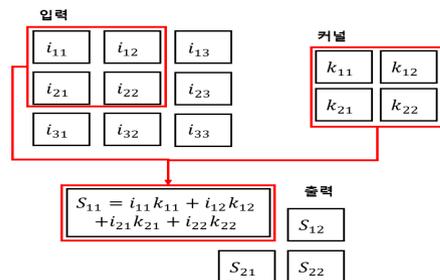


그림 1. 합성곱 계층의 기본 연산과정.
Fig. 1. Basic process of convolutional layer.

III. 실험 환경 및 결과

1. 실험 데이터

본 연구에서는 Leeds Sports Pose 데이터셋을 이용하였다^[7]. Leeds Sports Pose 데이터셋은 스포츠 동작을 취하고 있는 사람의 이미지 12,000 장으로 구성되어 있는 데이터셋이다. 또한, 이 데이터셋은 이미지에 대응하는 14개의 주요 관절의 x, y 좌표를 데이터 라벨로써 제공한다. 본 논문에서는 12,000 장의 데이터셋을 9,900 장의 학습데이터, 2,000 장의 검증데이터 그리고 100 장의 테스트 데이터로 나누어 실험을 진행하였다. 또한, 이미지 데이터는 임의의 각도 θ 만큼 회전시켜 학습과 검증을 진행하였다. 이때 임의의 회전각도 θ 는 범위로 $-45^\circ \leq \theta \leq 45^\circ$ 표현된다.

2. 실험 환경

본 연구에서 설정한 실험 파라미터는 총 4개이다. mini-batch size, 검증 주기, epoch 그리고 초기 학습률이다. 각 파라미터에 대해 간략히 서술하겠다.

먼저 mini-batch size이다. 이 파라미터는 모델이 학습을 할 때, 입력받아 연산할 데이터의 크기로, 이 값은 128로 정하였다. 다음으로 검증 주기는 검증 데이터를 통해 학습 방향을 검토하는 것으로, 본 실험에서는 검증 주기를 32로 설정하였다.

Epoch는 학습 데이터의 반복학습 횟수를 나타내는 파라미터이다. 본 연구에서는 이 파라미터를 100으로 설정했다.

마지막으로 본 논문에서는 인공지능 학습 optimizer 로써 Adam optimizer를 사용하였다^[8]. 이에 따른 초기 학습률은 0.001로 설정하였고, 20 epoch마다 학습률이 0.1배씩 더 낮아지게 설정하였다. 표 1은 본 연구에서 사용한 파라미터를 정리한 표이다.

표 1. 실험 파라미터.

Table 1. Parameters of experiments.

Parameters	Value
Mini-batch size	128
Validation frequency	32
Epoch	100
Initial learning rate	0.001

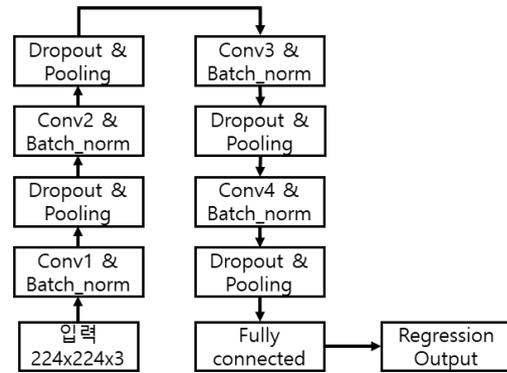


그림 2. CNN 네트워크.
 Fig. 2. CNN network.

3. CNN 네트워크

그림 2는 본 논문에서 사용한 CNN 네트워크를 보여준다. 네트워크의 입력으로는 가로 및 세로 길이 224 픽셀의 RGB 이미지의 픽셀값이 들어가게 된다. 이후 4개의 합성곱 계층, 배치 정규화 계층, 드롭아웃 및 풀링 계층을 통해 연산된다. 모든 합성곱 계층의 필터 사이즈는 가로 및 세로 길이 3픽셀로 지정하였고 활성화함수는 모두 relu 함수를 이용하였다. 합성곱 계층 연산을 진행하고 퍼셉트론으로 이루어진 fully connected 계층으로 값이 모여 regression 출력을 내게 된다. 이때의 출력은 이미지의 회전 각도값이 나타나게 된다.

4. 실험 결과

그림 3은 실험 결과의 예시 중 하나로, 원본 이미지와 회전된 이미지 그리고 출력으로 나온 추정된 회전 각도에 대해 복원한 이미지를 보여준다. 해당 이미지에 대한 임의의 회전각도는 31° 이며, 인공지능 모델이 추정한 회전각도는 31.3041° 로 나타났다. 테스트셋에 대한 성능평가를 진행하기 위해 본 논문에서는 평가지표로써 mean absolute error (MAE) 지표를 이용하였다^[9]. MAE의 수식적 표현은 아래와 같다.

$$MAE = \frac{|\hat{y}_1 - y_1| + \dots + |\hat{y}_n - y_n|}{n} \quad (3)$$

위의 수식에서 \hat{y}_n 은 n번째 예측결과 즉, 모델의 출력을 의미하고 y_n 은 n번째 테스트 데이터 정답 값을 의미한다. 본 논문에서 진행한 실험은 100개의 테스트 셋에

대해 MAE 기준 4.5951의 값을 얻었다. 이는 100개의 샘플에 대해서 평균적으로 4.5951도 차이가 난다는 것을 의미한다.

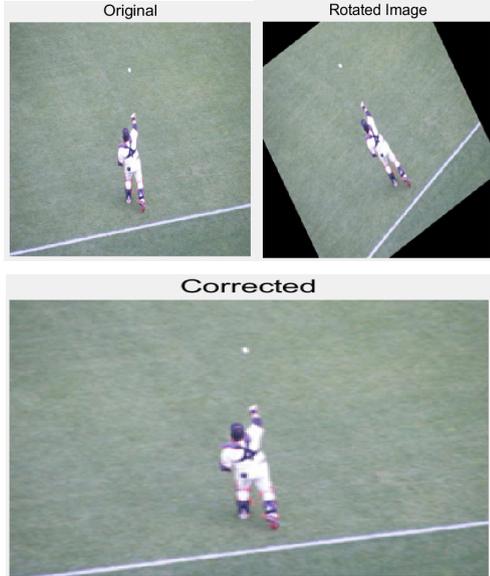


그림 3. 원본이미지(상단 좌측), 회전이미지(상단 우측), 회전 보정 이미지 출력(하단).
 Fig. 3. Original image(up left), rotated image(up right), corrected image(down).

IV. 결 론

본 논문에서는 Leeds Sports Pose 데이터셋을 활용하여 인공지능 기반 영상인식 시스템에서 인식을 저하 요인으로 나타나는 영상의 회전문제에 대해서 CNN 기반 regression 모델을 제안하였다. 제안한 모델을 학습시켜 임의의 회전 각도를 추정하는 실험을 진행하였고, 성능지표로써 MAE 지표를 통해 추정 오차 4.5951의 수치를 도출했다.

References

[1] I. Goodfellow, *Deep Learning*, MIT Press, 2016.
 [2] Y. LeCun, P. Haffner, L. Bottou and Y. Bengio, "Object recognition with gradient-based learning," *Proceedings of the Shape Contour and Grouping in Computer Vision Part of the Lecture Notes in Computer Science book series (LNCS)*, vol. 1681, pp.

319-345, 1999.

DOI: https://doi.org/10.1007/3-540-46805-6_19

[3] V. Mnih, et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529-533, Feb. 2015.
 DOI: <https://doi.org/10.1038/nature14236>
 [4] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Proc. of Neural Information and Processing Systems*, pp. 1106-1114, 2012
 DOI: 10.1061/(ASCE)GT.1943-5606.0001284
 [5] A. Bearman and C. Dong, "Human pose estimation and activity classification using convolutional neural networks," *CS231n Course Project Reports*, 2015.
 [6] S. P. Kim, *Deep Learning First Step*, Hanbit media, 2016.
 [7] S. Johnson and M. Everingham, "Learning effective human pose estimation from inaccurate annotation," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR2011)*, pp. 1465-1472, Providence, RI, USA, June 2011.
 DOI: 10.1109/CVPR.2011.5995318
 [8] D. P. Kingma and J. L. Ba, "ADAM: A method for stochastic optimization," *Proc. 3rd International Conference for Learning Representations*, pp. 1-41, San Diego, USA, 2015.
 [9] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp.1247-1250, June 2014.
 DOI: <https://doi.org/10.5194/gmd-7-1247-2014>

저 자 소 개

이 동 구(준회원)



- 2019년 2월 : 광운대학교 전자융합공학과 학사 졸업
- 2019년 3월~현재 : 광운대학교 전자융합공학과 석박사통합과정
- 관심분야 : 인공지능, 위성통신, 스마트그리드

선 영 규(준회원)



- 2018년 2월 : 광운대학교 전자융합공학 학사 졸업
- 2018년 3월~현재 : 광운대학교 전파공학과 석박사통합과정
- 관심분야 : 인공지능, 무선 에너지하베스팅, 디지털통신, 전력선 통신

송 명 남(준회원)



- 2001년 2월 : 서울여자대학교 컴퓨터학과 석사
- 2001년 ~ 2012년 : (주)XTM 월드 전자팀 차장
- 2012년 ~ 현재 : (주)현대엠아이비인터내셔널 기술개발팀 팀장
- 관심분야 : 영상처리, 신호처리, 임베디드 시스템

김 수 현(준회원)



- 2019년 2월 : 광운대학교 전자융합공학과 학사 졸업
- 2019년 3월~현재 : 광운대학교 전자융합공학과 석박사통합과정
- 관심분야 : 인공지능, 전력선통신, 스마트 그리드

김 진 영(정회원)



- 1998년 2월 : 서울대학교 전자공학과 공학박사
- 2001년 2월 : SK텔레콤 네트워크 연구소 책임연구원
- 2001년 3월~현재 : 광운대학교 전자융합공학과 교수
- 관심분야 : 차세대이동통신, 가시광통신, 전력선통신, 인공지능

심 이 삭(준회원)



- 2016년 2월 : 광운대학교 전자융합공학과 졸업
- 2016년 3월~ 현재 : 광운대학교 전파공학과 석박사통합과정
- 관심분야 : 전력선 통신, 무선 에너지 하베스팅, Backscatter, 인공지능

이 계 산(정회원)



- 2002년 2월: 게이오대학교 전자공학과 공학박사
- 2002년~2003년 : 일본 KDDI 연구원
- 2003년~ 현재: 경희대학교 전파공학과 교수
- 관심분야 : DS-CDMA, OFDM, LED-ID, LTE-A

※ 위 논문은 문화체육관광부의 스포츠산업기술개발사업에 의해 국민체육진흥공단의 국민체육진흥기금을 지원받아 연구되었습니다.