

중첩신경망에서 그라디언트상승최적화에 의한 중간층특징의 시각화방법

곽 혁

경애하는 김정은동지께서는 다음과 같이 말씀하시였다.

《지식경제시대인 오늘 과학기술은 사회경제발전을 규제하고 떠밀어나가는 결정적요인으로 되고있습니다.》

심층학습은 화상인식, 자연언어처리를 비롯한 다양한 분야에서 혁신적인 성능개선을 가져온다.

그러나 심층학습은 동작원리에 대한 수학적해명이 부족하며 성능이 개선되는 리유조차도 충분히 설명하지 못하고있다. 특히 신경망의 구조와 학습에 필요한 하이퍼파라미터를 어떻게 설정할것인가에 대한 이론적인 해명이 밝혀진것이 없으며 좋은 결과를 얻었다고 하여도 그것을 수학적으로 론증하는것이 어렵다.

이로부터 많은 연구자들이 신경망을 보다 구체적으로 이해하고 학습결과를 확증하기 위한 방법[1, 2]들을 제안하였다.

중첩신경망의 시각화기술은 중간층의 출력이나 학습파라미터, 신경망의 최종출력결과를 화상이나 그래프로 시각화하는 기술로서 신경망의 구조를 이해하는데 효과적인 방법으로 된다.

논문에서는 중첩신경망에서 중간층출력인 특징지도(feature map)를 시각화하는 한가지 방법을 제안하였다.

1. 중간층출력의 시각화를 위한 그라디언트상승최적화

중첩신경망에서 중첩층과 같은 중간층들의 출력결과는 크기가 $w \times h \times m$ (w, h 는 특징지도의 가로, 세로방향크기, m 은 특징의 개수)인 3차원배렬형태의 특징지도이다. 이것을 화상으로 시각화하면 해당 층의 특성과 동작원리를 파악할수 있으며 신경망의 구조설계와 하이퍼파라미터들을 설정하는데도 활용할수 있다.

선행연구[1]에서는 AlexNet의 5번째 응집층에서 출력값이 최대로 되는 입력화상이 감수영역(receptive field)과 밀접한 관계가 있다는것을 보여주었다. 또한 회전, 자르기, 크기변화 등 다양한 변형을 가하여 생성한 입력화상들중에서 유니트의 출력값이 최대인 화상들이 해당 유니트가 원본화상에서 어느 부분에 주목하는가를 나타내는 감수영역에 해당하며 유니트의 기능과 밀접한 관계에 있다는것을 실험적으로 확인하였다. 이것은 이 화상들이 중간층유니트의 출력을 시각화하는 대표화상으로 될수 있다는것을 암시해준다.

그러나 표본화상의 종류와 개수를 선정하는 기준이 명백하지 않은것을 비롯하여 여러가지 문제점들이 있으므로 학습자료기지나 시험자료기지에서 대표화상을 수동적으로 찾아내는것은 적합하지 않다.

이로부터 논문에서는 중간층특징의 시각화를 중간층출력을 최대화하는 입력화상을 찾는 최적화문제로 일반화하여 해결하는 한가지 방안을 제안하였다.

론문에서 제안한 방법을 다음과 같이 정식화할수 있다.

$$x^* = \arg \max h_{ij}(\theta, x) \quad (1)$$

여기서 θ 는 중첩신경망의 학습파라미터(무게와 편위), h_{ij} 는 신경망의 j 번째 층 i 번째 유니트의 출력이다.

h_{ij} 는 신경망파라미터 θ 와 입력화상 x 의 함수이지만 이미 학습된 신경망모형인 경우에는 θ 를 고정하고 x 만의 함수로 볼수 있다.

식 (1)은 대역최대점이 아닌 국부최대점을 찾는 비볼록성(non-convex)최적화문제로서 그라디언트상승법을 리용할수 있다.

$$x \leftarrow x + \alpha \cdot \frac{\partial h_{ij}(x, \theta)}{\partial x} \quad (2)$$

식 (2)는 일반적인 신경망의 최적화알고리즘인 그라디언트하강법과 유사하게 학습률(α)과 반복회수와 같은 하이퍼파라미터를 포함한다. 그리고 h_{ij} 는 일반적으로 ReLU활성화함수를 적용하기 전의 출력값으로 한다. 그것은 부의 값을 가지는 특징값들의 그라디언트를 영이 아닌 값으로 보존하기 위해서이다.

2. 실험 결과

우리는 매물형장치용얼굴인식신경망의 설계와 학습에 제안한 방법을 적용하였다.

론문에서 제안한 방법을 심층학습프레임워크인 Tensorflow를 리용하여 코드로 구현하였다.

식 (2)에서 x 의 초기값은 신경망모형의 구조에 맞게 정규화한 얼굴화상으로, 학습률(α)은 1.0, 반복회수는 20으로 설정하였다.

우선 학습완료된 얼굴인식신경망모형에 대하여 첫 단계와 중간단계, 마지막단계의 중첩층에서 대표적인 5개 통로의 출력특징지도를 시각화하였다.(그림 1, 2, 3)



그림 1. 첫 단계 중첩층에서 특징지도의 시각화결과

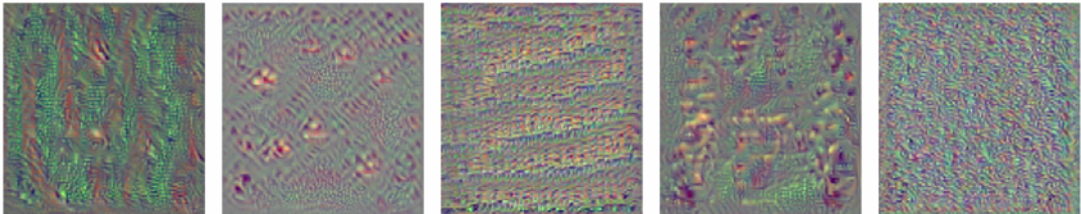


그림 2. 중간단계 중첩층에서 특징지도의 시각화결과

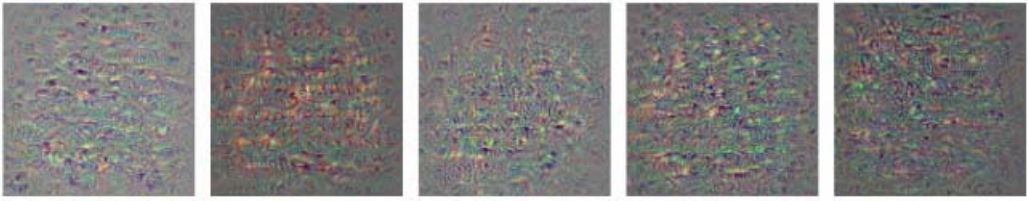


그림 3. 마지막단계 중첩층에서 특징지도의 시각화결과

시각화결과를 분석해보면 첫 단계에서는 색, 무늬와 같은 일반적인 저준위특징을 추출하고 마지막단계에서는 얼굴주위에서 규칙적인 패턴모양이 나타나면서 해당 얼굴에 고유한 고준위특징을 추출한다는것을 확인할수 있다. 그리고 특징이 어느 한 려파기(통로)에 집중되지 않고 각 통로마다 고유한 특징이 나타나고있다.

다음으로 중첩층의 깊이별로 특징을 포착하기 시작하는 학습반복회수를 측정한 결과 첫 단계 중첩층에서는 20회, 중간단계에서는 100회, 마지막단계에서는 200회로서 각 층별로 학습되는 시간이 다르며 마지막단계의 중첩층에서 특징이 나타나는 반복회수를 신경망의 학습반복회수로 정할수 있다는것을 확인하였다.

각이한 조건에서 촬영한 동일한 인물의 얼굴화상을 초기값으로 설정하고 첫 단계와 마지막단계에서 중첩층의 특징지도를 시각화한 결과는 그림 4와 같다.

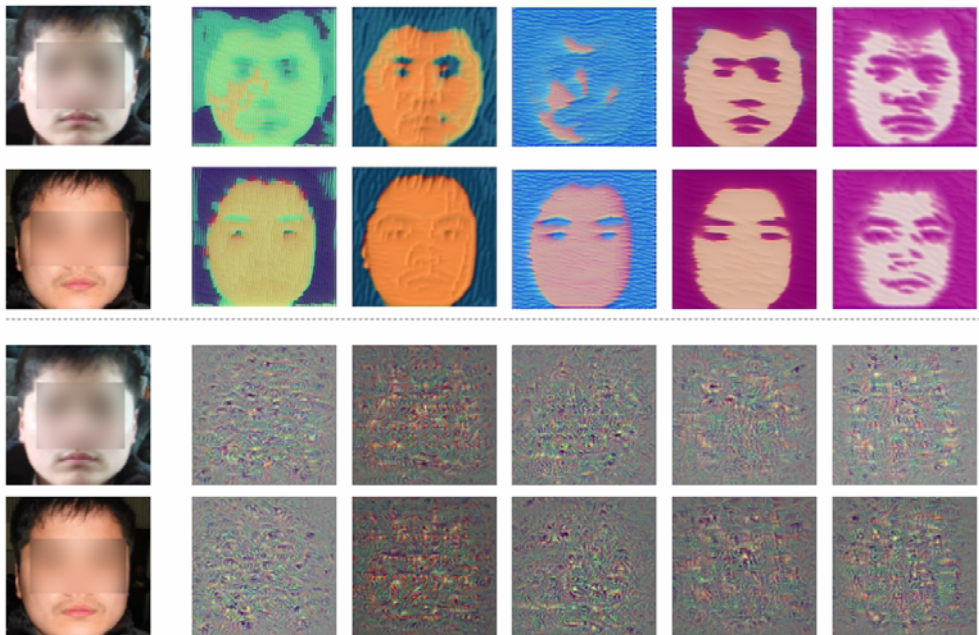


그림 4. 동일한 인물의 각이한 얼굴화상에 대한 시각화결과

왼쪽 첫렬은 입력한 얼굴화상이며 우부분은 첫 단계, 아래부분은 마지막단계 중첩층의 특징지도를 시각화한것이다.

그림 4로부터 확인할수 있는바와 같이 첫 단계 중첩층에서는 얼굴변화에 대하여 눈에 띄울 정도로 변화가 있는 특징지도가 출력되지만 마지막단계의 중첩층에서는 얼굴변화에 의존하지 않는 유사한 형태의 특징지도가 출력된다는것을 알수 있다.

맺 는 말

그라디언트상승최적화에 의한 중첩신경망의 특징시각화방법을 제안하고 매물형장치용얼굴인식신경망을 학습하는데 적용하여 그 효과성을 확인하였다.

참 고 문 헌

- [1] R. Girshick et al.; Computer Vision and Pattern Recognition, 4, 7, 580, 2014.
- [2] M. D. Zeiler, R. Fergus; Computer Vision—ECCV 2014, Springer, 818~833, 2014.

주체110(2021)년 5월 5일 원고접수

Visualization of the Feature Maps of Convolutional Networks Using Gradient Ascent Optimization

Kwak Hyok

This paper accounts for the visualization of the feature maps of the convolutional networks using gradient ascent optimization algorithm and applies it to training the efficient CNN models for face recognition on mobile and embedded devices.

Keywords: visualizing convolutional networks, gradient ascent optimization