

AlexNet 논문 공부

- 2012년 이미지 분해 대회에서 우승을 차지한 모델
- CNN을 Feedforward Neural Network 보다 더 적은 파라미터와 복잡성이 있어 학습하기 쉬움으로 보고, 이론상 초에 성능은 Feedforward Neural Network 보다 약간만 적어졌다.
- dataset의 형식, 고정된 크기의 이미지가 필요한 256x256으로 이미지 크기를 줄이고, 정사각형 이미지를 만들기 위해 이미지를 crop하였다. 그리고 각 픽셀의 RGB 값을 전체 이미지의 RGB 공간으로 해서 이미지 레이어를 centering 하였다.
* centering은 각 픽셀의 픽셀값을 0 주변으로 끌어들이고서, 초점화 알고리즘이 중심 합들의 차이값을 학습하게 더 빠르게 도달할 수 있다.

★ Architecture

1) ReLU Nonlinearity

- saturating nonlinearities를 사용하는 것보다 non-saturating nonlinearities를 사용하는 것이 학습 에러를 낮추는데 걸리는 시간이 6배 감소된다고 한다. saturating nonlinearities의 경우 미분값이 작아지면 결국 vanishing gradient 문제가 발생하게 된다.

2) Multiple GPUs

- 특정 데이터셋에 대해 GPU를 병행적으로 사용한다.

3) Local Response Normalization

- ReLU는 양의 방향으로만 입력의 값을 그대로 사용하기 때문에, conv나 pooling 시 마다 높은 하나의 픽셀값이 주변이 픽셀에 영향을 미치게 된다. 이런 문제를 방지하기 위해 Activation Map의 같은 위치가 있는 픽셀끼리 정규화를 하였다.
- 깊은 레이어에서는 LRN의 효과가 더 중요하기 때문에 정규화는 처음 몇몇 레이어에만 적용했으며, ReLU 이후 강한 활성화된 뉴런들을 균형 있게 조정하기 위해 ReLU를 거칠고 난 뒤라값을 사용하였다.

$$b_{i,y} = a_{i,y} / \left(k + \alpha \sum_{j=\max(0, i-\frac{n}{2})}^{\min(N-1, i+\frac{n}{2})} (a_{j,y})^2 \right)^\beta$$

정규화된
활성화 값

원래
활성화 값

feature map의 원리 뉴런과
min(N-1, i+n/2)
j=max(0, i-n/2)
↑
주변 뉴런의 개수 n과,
활성화 값을 정규화하기
위해 사용하는 이웃 뉴런의 수

k, α, β : 하이퍼파라미터

- 논문에서는 k=2, n=5, α=10⁻⁴, β=0.75로 설정하였다.
- 광범위한 지역 뉴런이 주변 뉴런들에 비해 매우 강하게 활성화된다면, 해당 뉴런의 활성화 값을 줄여주고, 주변 뉴런들에 비해 활성화가 약하면, 해당 뉴런의 활성화 값을 그대로 둔다.

4) Overlapping Pooling

- 겹쳐서 pooling 원상 복원. 겹치는 영역으로 인해 같은 특징이 여러 pooling 윈도우에 의해 캡처될 수 있고, 관련 윈도우의 여러 특징에 큰 영향을 미치지 않게 만들어, 모델의 robustness를 향상시키는 효과가 있다.
- stride=2, kernel=3x3 사용

t) Overall Architecture

- 총 8개의 레이어로, 5개는 Convolution Network, 3개는 Fully Connected Network.
- 파라미터 개수는 약 6000만 개

★ Reducing Overfitting

1) Data Augmentation

- 회전, 뒤집기, 패치 사용
- RGB 채널의 색상강도 조정 (Color PCA)

2) Dropout

- 사용자가 지정한 확률을 근거로 하여 특정 뉴런에 신호를 전달하지 않는 방법
- 1번째와 2번째 fc layer에 50% 확률로 Dropout 적용