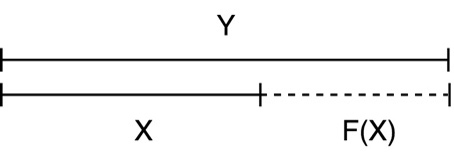
Image classification tutorial

Jo MinKi

1. Introduction  
   Kakao 인턴의 따스한 온기가 느껴지는 문서이다. 프로젝트는 검색 엔진에 올려질 유해 이미지 필터를 위한 분류기 구현으로, 기본적으로는 이미지 분류 튜토리얼 정도가 될 것 같다. 해결해야 할 문제는 몇가지 고려되어야 할 특징들이 있다. 1) 포털 서비스에 업로드 될 이미지에는 압도적으로 비 유해이미지가 많기 때문에 입력되는 이미지들을 그대로 인풋으로 주며 Accuracy를 측정하는 것은 의미가 없다. 2) 음란물, 혐오 이미지들을 최대한 필터링 해야 하며, 검색 엔진이 현존하는 모든 웹사이트를 탐색하여 결과를 출력하는 것이 아니기 때문에 Precision을 높게 잡는 것이 좋다.
2. Configuration  
   학습데이터는 총 300Gb의 학습 이미지셋이 2개, 30Gb의 이미지 셋이 하나 포함되어있다. 타 부서에서 분류된 106만개의 이미지가 들어 있으며 ADULT, SEMI\_ADULT, HATRED, AD, NORMAL이렇게 6가지 카테고리로 분류되어 있다. 할당된 GPU는 T-40 4개, CPU 16코어 및 메모리 96GB이다.
3. Main concept of this is that instead of target function , train the network to fit to the residual function . For example, if the target function is identity function , then training for instead of . This concept could be treated as a shortcut connection, which means connect one or more layers without any parameter.   
   Basic concept of ResNet is based on VGG Net, so the filer has 3x3 fixed size filter. And to make the network gets deeper, the researchers doubled the number of filter when the network has a pooling.
4. Deep Residual Learning   
   (I don’t get this part really much so I’m just stating their logic as they claim without any doubt. But my predecessor’s thinking is that making some plausible excuses is needed for researchers even though themselves don’t agree with that. So I think we can just take this claim like ‘Ah~ that could be the case’.)  
     
   

Question.

1. Pretrained model을 사용해서 180 이미지용 필터를 만들어 내는 방법?  
   A. 제공된 모델 내부의 Max pooling 부분을 이미지 규격에 맞게 조절해주면 됨
2. Gpu가 하나만 돌아가는 이유가? -> nn.DataParallel(model) 로 해결
3. 모델을 저장해 뒀더니 후미의 fc와 dataparellel까지 다같이 저장이 되었음. 뒷부분을 해제하고 앞부분 파라미터만 사용할 수 있는 방법이?
4. Criterion에서 0으로 깎아서 쓰지 말고 smoothing 하는 방법?
5. Precision 조절하는 방법?
6. 이 프로젝트 수주 받을때 구체적인 주문 내용? 혹은 이 서비스가 어디다 쓰이는지? (구체적으로는 SEMI\_ADULT 클래스의 존재가 필요한지?
7. 클래스랑 레이블 매핑이 된걸 어떻게 확인하는지? 보기로는 할 때 마다 랜덤으로 다시 매핑되는 것으로 예상됨.