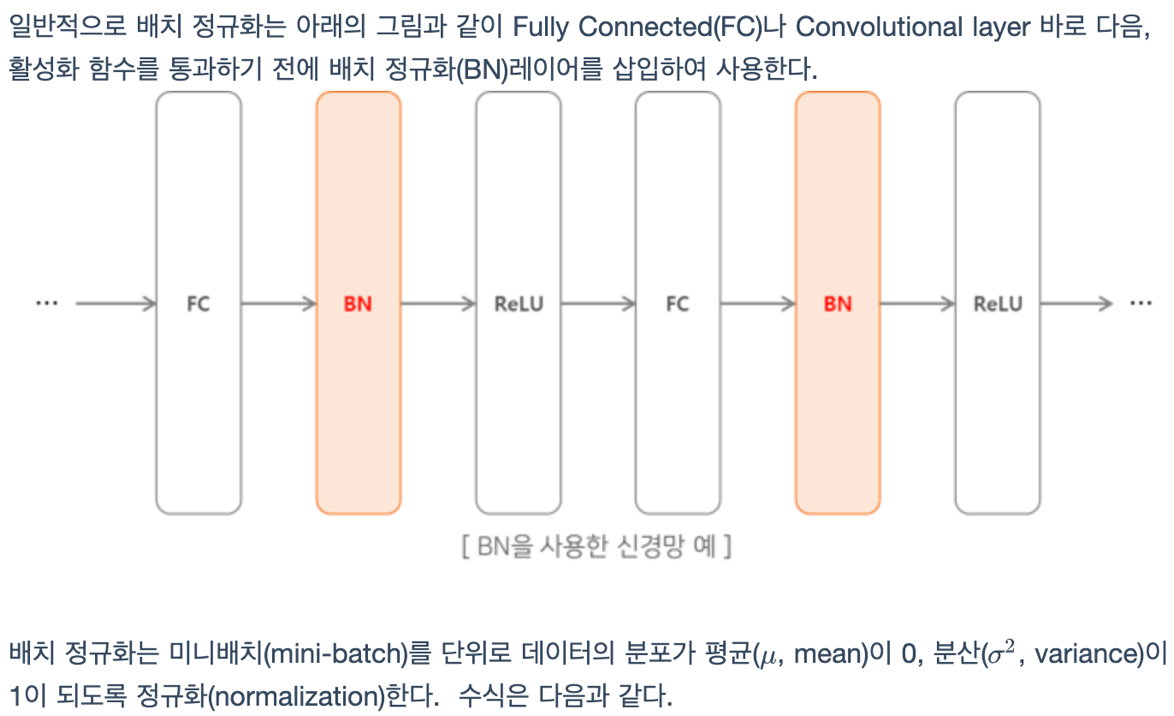
*From* [*https://excelsior-cjh.tistory.com/178*](https://excelsior-cjh.tistory.com/178)

ReLU + He initialization -> 학습 초기 단계에서 vanishing & exploding gradient 발생 가능성 존재

1. **Batch normalization**

각 층의 활성화 함수의 출력값 분포가 골고루 분포되도록 강제하는 방법

1. Mini-batch data에서 feature별 mean, variance 구한 후 normalize 하기



텍스트이(가) 표시된 사진

자동 생성된 설명

Scale & shift

Normalize 결과(x\_hat)가 0에 근사하면 sigmoid가 선형성을 보임

Scale\_parameter=0, shift\_parameter=0으로 초기화한 후 back propagation 과정에서 조정

전체 train data가 너무 크므로 n개의 mini-batch에 대한 평균과 표준편차로 추정

이것 대신 exponential decay / moving average 사용하여 mean, std 구할 수 있음

텍스트이(가) 표시된 사진

자동 생성된 설명

장점

1. Activation function에 대한 vanishing gradient 나타날 가능성 감소
2. Weight initiation에 덜 의존적
3. Learning rate 크게 잡아도 gradient descent 잘 수렴함
4. Overfitting 막는데 도움

텍스트이(가) 표시된 사진

자동 생성된 설명

1. **Gradient clipping**

Exploding gradient 가능성 줄이는 방법

