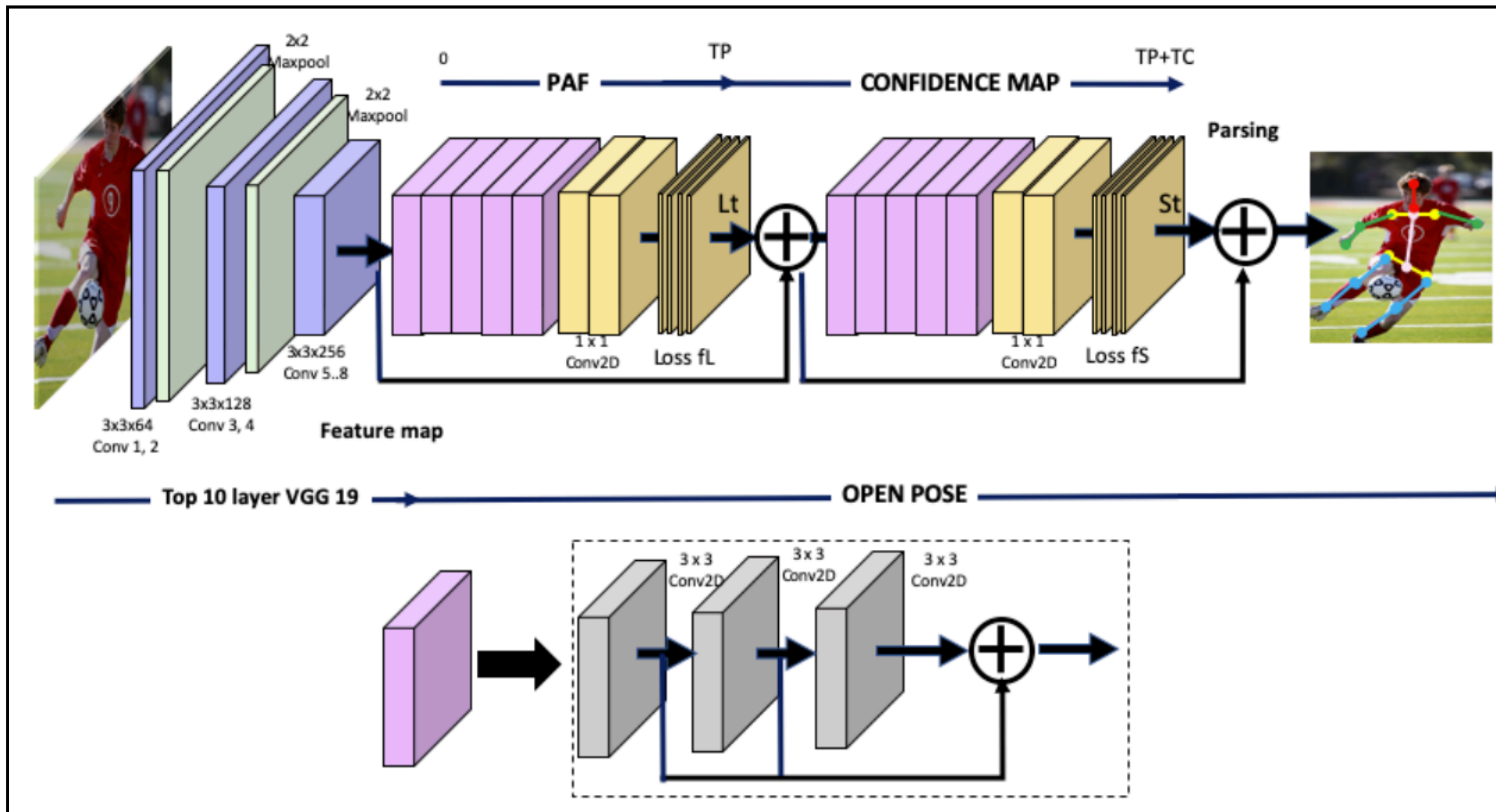


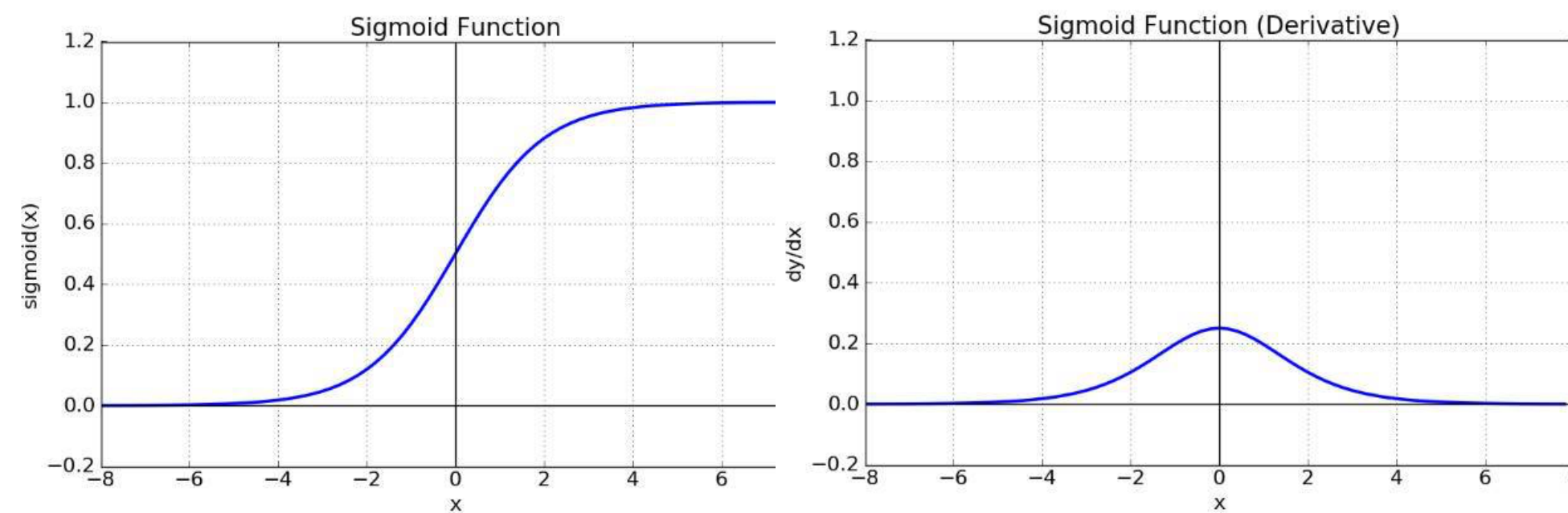
## Network

- ▶ C = Convolution / P = Pooling / S = Confidence Map / L = Part Affinity Field / F = Feature Map
  - Stage가 여러 개인 이유는 네트워크가 깊어질수록 앞에 위치한 Layer는 학습의 영향이 줄어들음.
    - ▶ 1개의 Stage보다는 여러개의 Stage를 사용하여 중간 Stage 사이에 결과값을 특징맵(F)로 만들어 입력하는 구조를 이용



# Degradation Problem: Vanishing Gradient

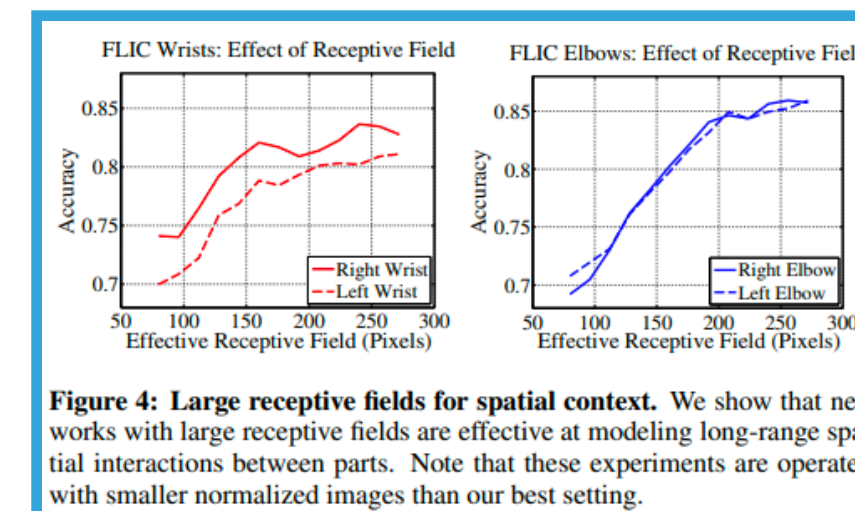
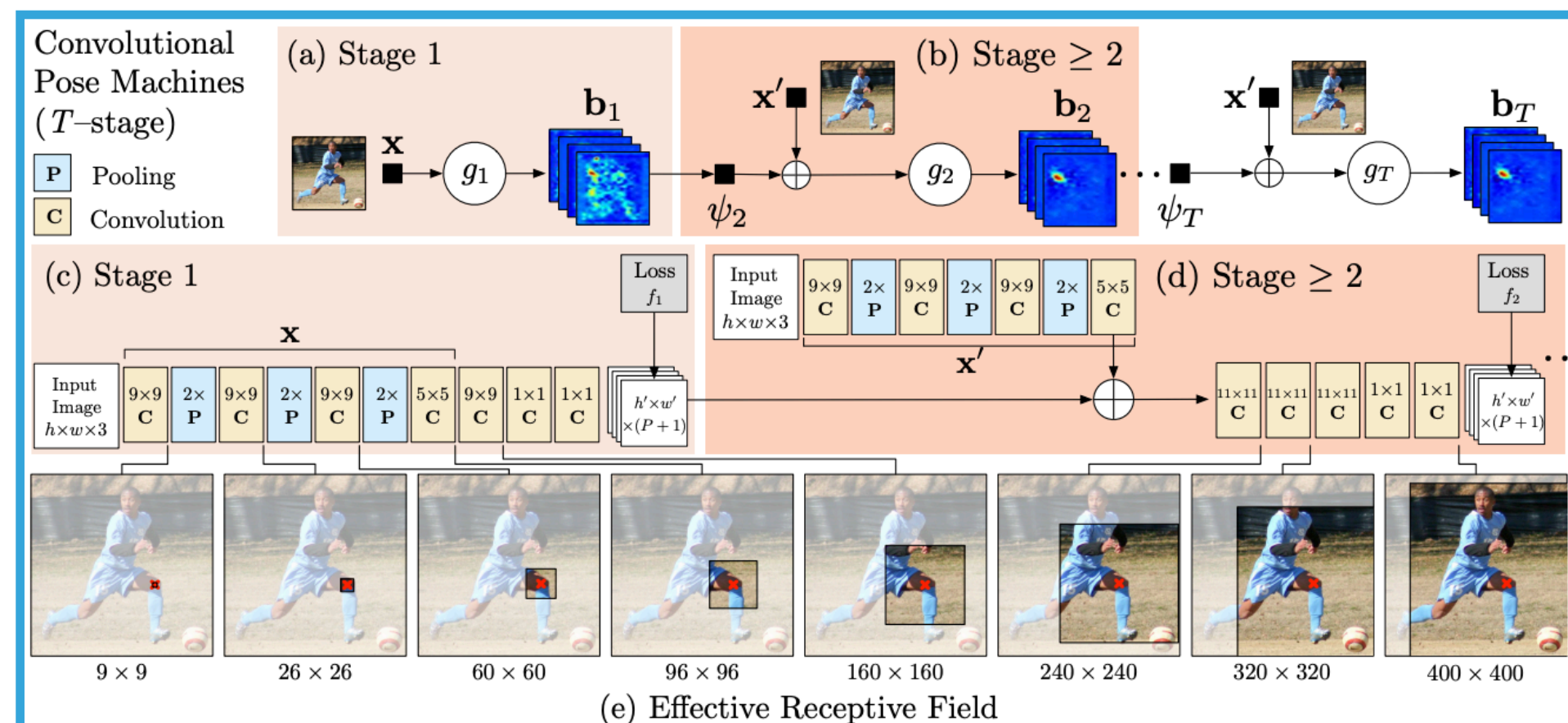
- ▶ 기울기 소실(Gradient Vanishing): 깊은 인공 신경망을 학습할 때, 역전파 과정에서 입력층으로 갈수록 Gradient가 점차적으로 작아지는 현상이 발생. 입력층에 가까운 층들에서 가중치들이 업데이트가 제대로 되지 않으면서 최적의 모델을 찾을 수 없게 됨.
  - 기울기 소실 예시) 시그모이드 함수



- ▶ 학습을 할수록 기울기가 작아져서 0에 가까워짐. (학습이 잘 안됨)

# Region, Global - Contextual

- ▶ CPM: 큰 receptive field를 이용 because of the problem of small receptive field
- ▶ CNN을 이용 - 하위 레이어에서 local한 영역을 해석하고, 상위 레이어로 갈수록 receptive field가 커지면서 global한 영역을 해석할 수 있게 됨.
  - local한 영역에서 global한 영역으로 확대하여 다른 부위와의 관계 고려함
  - receptive field를 키우고, context 정보를 이용하여 joint의 추정 정확도를 향상시킴
- ▶ Stage > 1의 input (그림 (b))
  - heatmaps and image evidence(이전 단계로부터의 context 정보) as input
  - input heatmaps add spatial context for the next stage
- ▶ 특정 관절의 heatmap 생성 시, 다른 파트의 heatmap 참조(spatial context 제공)를 통해 어려운 part를 쉽게 찾을 수 있음

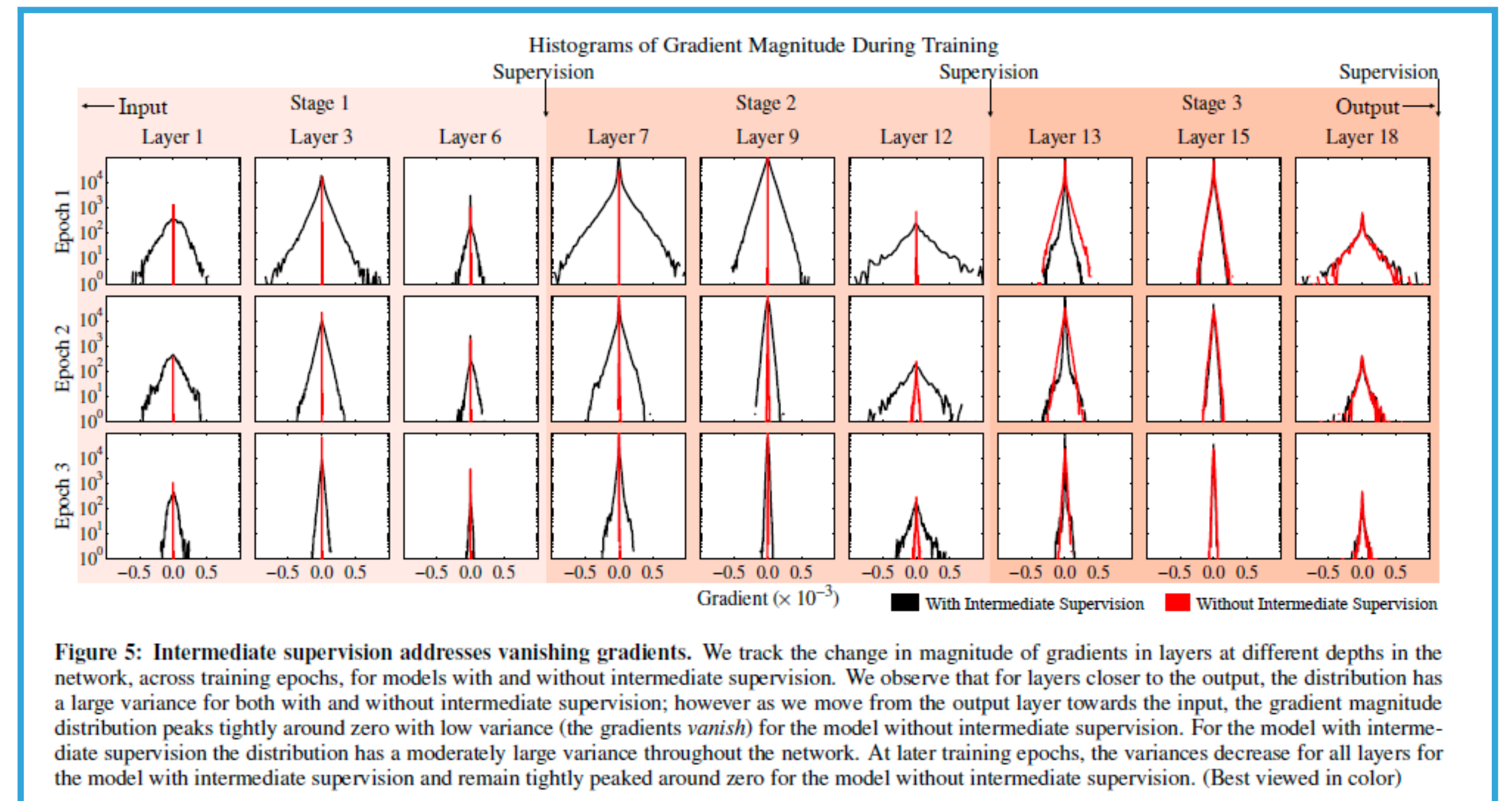




# Intermediate Supervision

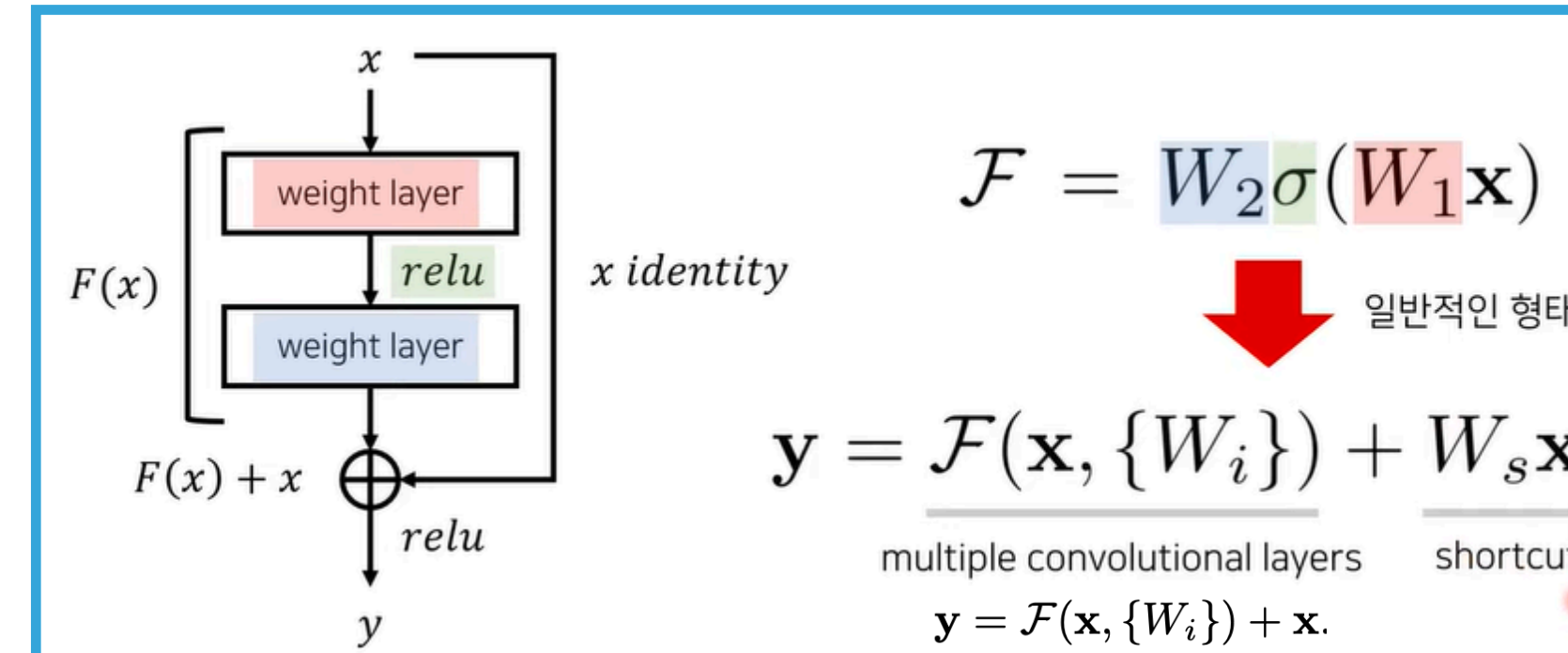
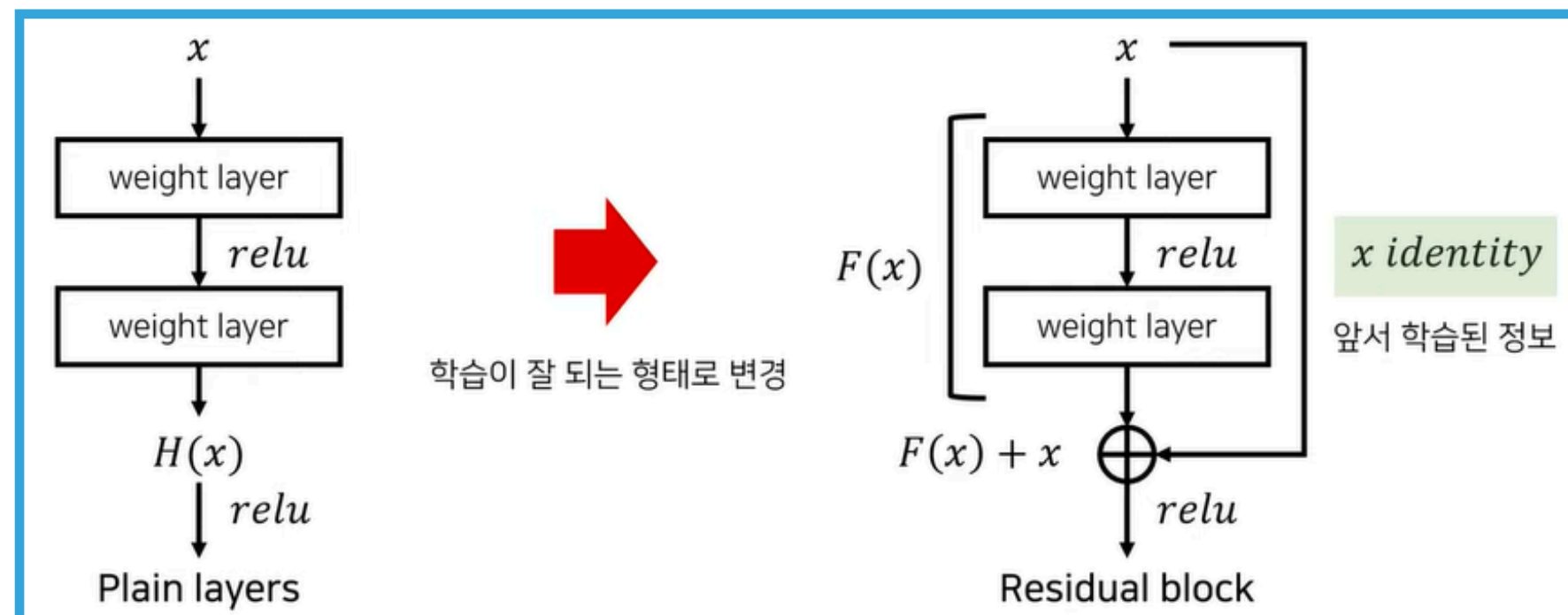
- ▶ Loss 함수를 the end of each stage 마다 배치 => gradient vanishing problem 해소
- ▶ 빨간색 그래프(without)의 폭은 매우 좁거나 0에 가까움. => 기울기 소멸문제로 학습이 중단된 것을 의미
- ▶ 검은색 그래프(with)의 폭은 꽤 넓음 => 기울기가 안정적으로 소멸하는 것으로 제대로 학습이 되고 있음을 의미
- ▶ 해결방식

- the intermediate loss functions replenish the gradients at each stage
- all layers의 기울기 much larger variance across all layers => all layers에서 학습이 잘 이루어질 수 있음
- **How?**



# Residual Learning 잔여 학습

- ▶ 기존의 CNN 모델(VGG Net)은 레이어가 깊어질수록 모델의 학습 난이도가 높아지며, 의도대로 최적화 시키는 것이 어려움
- ▶ 깊은 네트워크 학습하기 위한 방법
  - 레이어가 깊어질수록 성능이 올라감 (예상)
  - 레이어가 너무 깊어지면 오히려 성능이 떨어짐 by degradation problem
- ▶ Residual Block 잔여 블록
  - Input은 그대로 가져오고, 나머지 잔여 정보인  $F(x)$ 만 추가적으로 더해줌.
  - 잔여 정보인  $F(x)$ 만 학습해도 OK.  $H(x)$ 보다 학습이 쉬어짐. (네트워크의 optimization 난이도를 낮춤)





## ResNet

- ▶ residual block을 활용하면, layer가 깊어질수록 error가 줄어듬
  - Identity Mappings: 자기자신을 더함
- ▶ (비교) DenseNet의 concatenation: feature들을 모두 합침
  - 이전의 모든 layer들로부터 추가적인 정보를 얻고, 만든 feature map은 그 다음 모든 layer로 전달
  - 모든 layer를 직접 연결 (정보를 최대한 많이 전달)
- ▶ Openpose: input들 concat ? (submission과 다름)

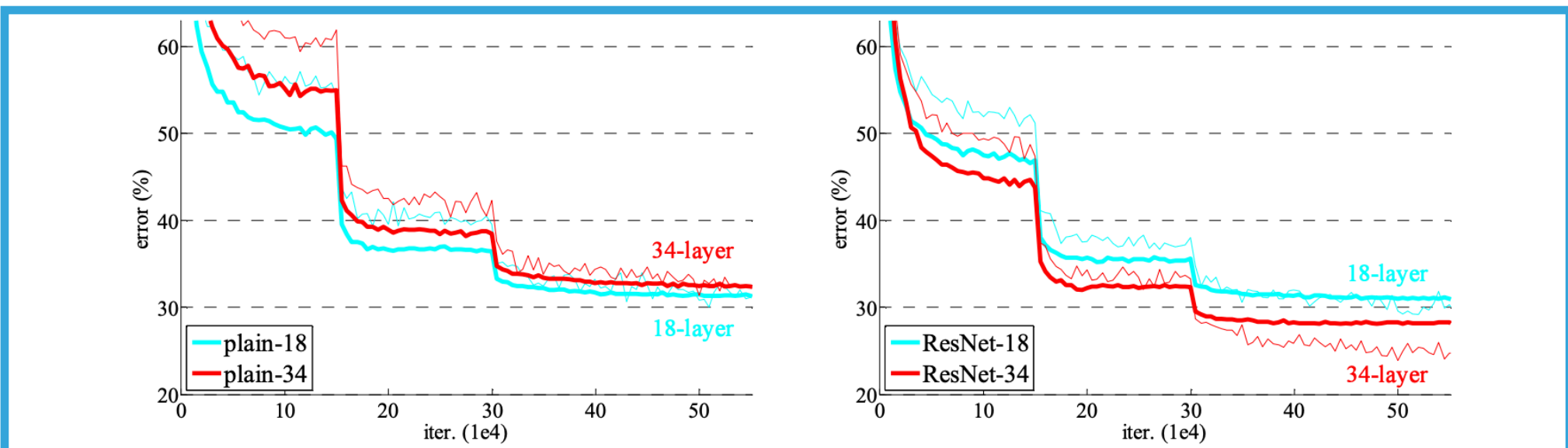
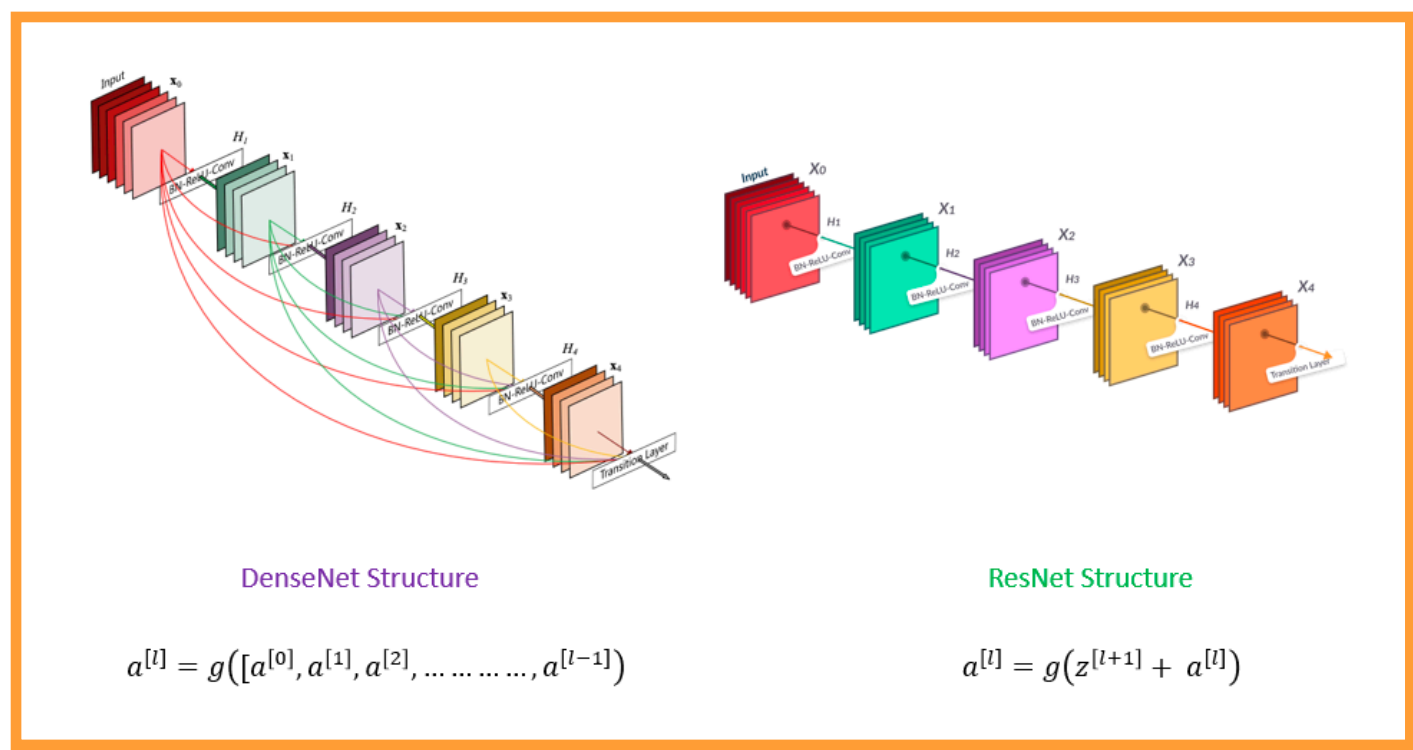
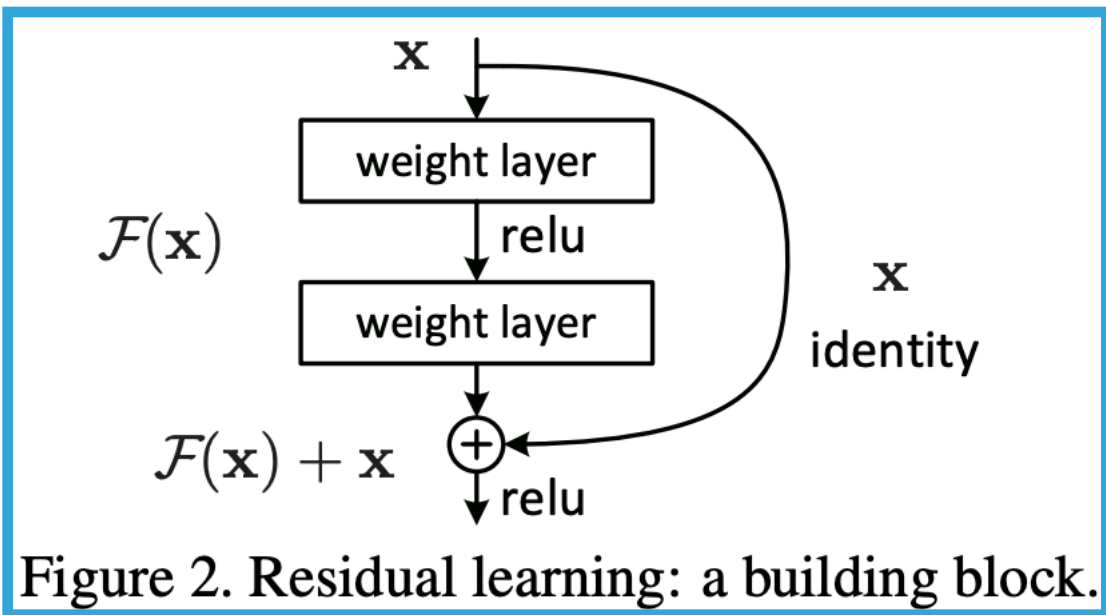


Figure 4. Training on **ImageNet**. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

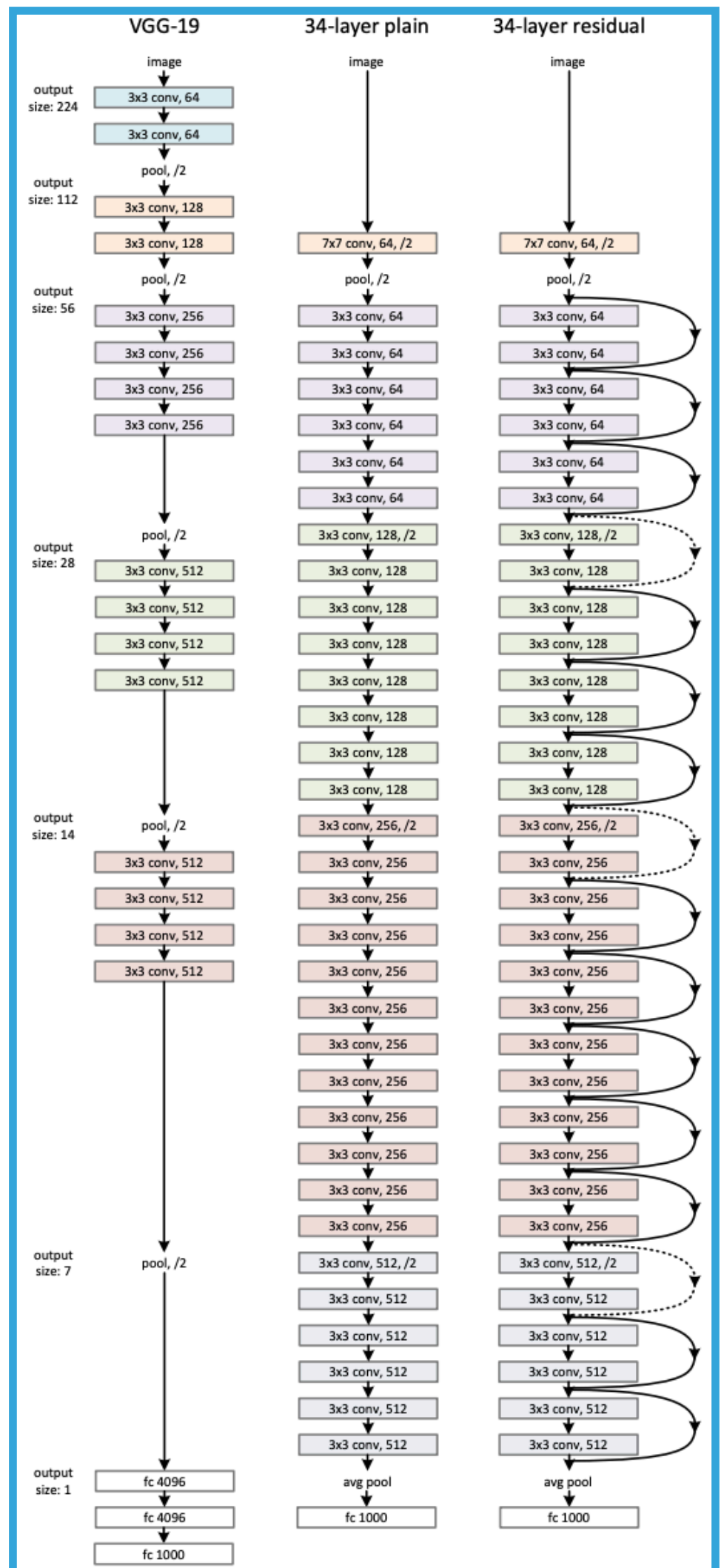


Figure 3. Example network architectures for ImageNet. **Left:** the VGG-19 model [41] (19.6 billion FLOPs) as a reference. **Middle:** a plain network with 34 parameter layers (3.6 billion FLOPs). **Right:** a residual network with 34 parameter layers (3.6 billion FLOPs). The dotted shortcuts increase dimensions. **Table 1** shows more details and other variants.