

■ Bag of Tricks for Image Classification with CNN

- 실험 네트워크는 ResNet-50
- AlexNet이후로 많은 네트워크가 나왔고, 성능도 꾸준히 올라가는 상황이다.
- Loss Function, Data preprocessing, Optimization 등의 변경으로 성능 증가가 이루어짐.
- 여러 가지 기법들을 소개하는 논문

■ Large-Scale Batch Training

- convex 문제에서? batchsize 클수록 가중치 수렴속도가 감소 → 이를 해결하기 위해?

1) Linear Scaling Learning rate

- Mini-batch SGD(Stochastic Gradient Descent) : Batch size 증가 → learning rate 증가가 좋다.

2) Learning rate Warmup

- 처음 learning rate는 아주 작은값에서 시작하고, 점진적으로(선형적으로) learning rate를 증가시키자.(~5 epoch 까지)

3) Zero Gamma

4) No bias Decay

- Overfitting 방지를 위해서 권장.
- Conv/FC Layer에만 적용

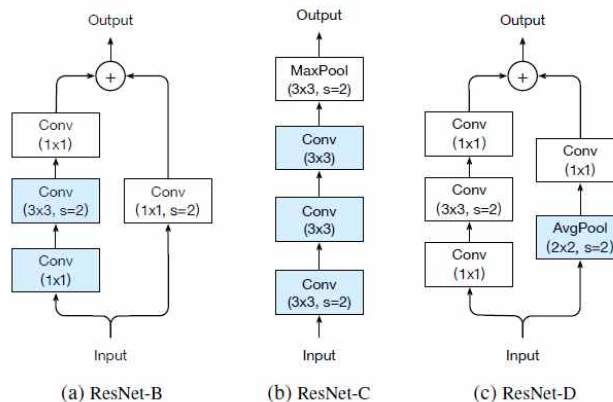
■ Low-Precision Training

- 작은 크기의 데이터 타입을 사용할 경우 수치 연산 향상을 얻을 수 있음(FP16 vs FP32).
- FP32에서 FP16으로 precision을 줄이면 → 성능 저하가 있지만, 2~3배 빠른 학습 속도.
- 여기의 성능 저하는 Mixed Precision Training을 적용하여 학습시킴.

- Nvidia V100이 FP16 지원, FP16으로만 구성된 모델 작성하고 FP32로 값 복사 후 loss 계산.

■ Model Tweaks

- 모델의 파라미터 등 변경, 복잡도 역시 변경 → 성능 향상이 이루어질 수 있다.
- ResNet-A(기본 모델)에서 Tweak을 적용한 B, C 모델 존재함, 추가적으로 D 모델 개발.



■ 정확도를 올리기 위한 4가지 방법(Training Refinement)

1) Cosine Learning Rate Decay

- learning rate는 학습 시 중요한 요소 중 하나.
- lr을 초기값으로부터 어떻게 감소시켜야 하나?
- exponentially decay, step decay에 이어서 cosine decay 소개
- 초반에는 천천히 감소, 후반부에는 급격히 감소

2) Label Smoothing

3) Knowledge Distillation

- 먼저 학습된 모델의 결과를 새로 배울 모델에게 가르쳐주면 성능이 향상됨.
- 모델의 복잡도를 줄이는데 사용.

4) Mixup Training

- 학습 할 이미지 2장을 섞는 기법

■ Transfer Learning 실험

- Object Detection은 VGGNet-19, Semantic Segmentation은 FCN 사용
- 적용 후, 성능이 향상됨.