

Swe3050

Fundamentals of machine learning

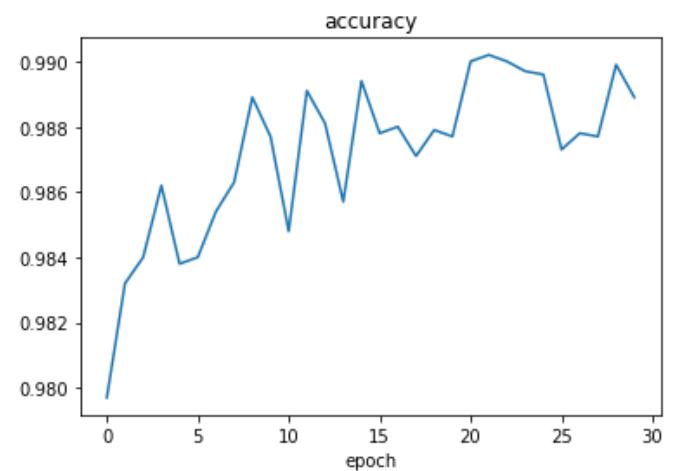
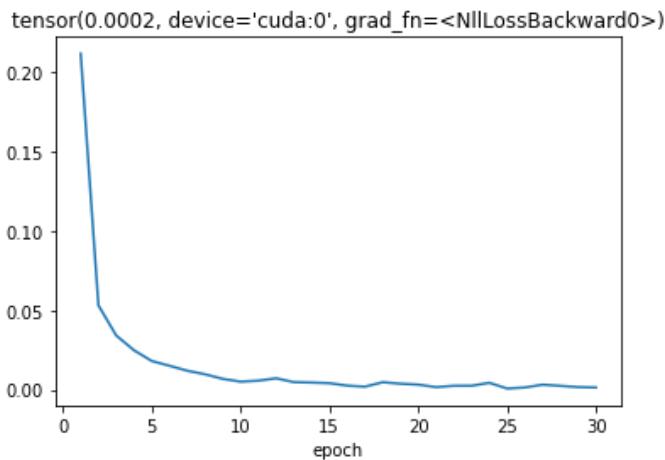
HW2

소프트웨어학과 2019315505

이원규

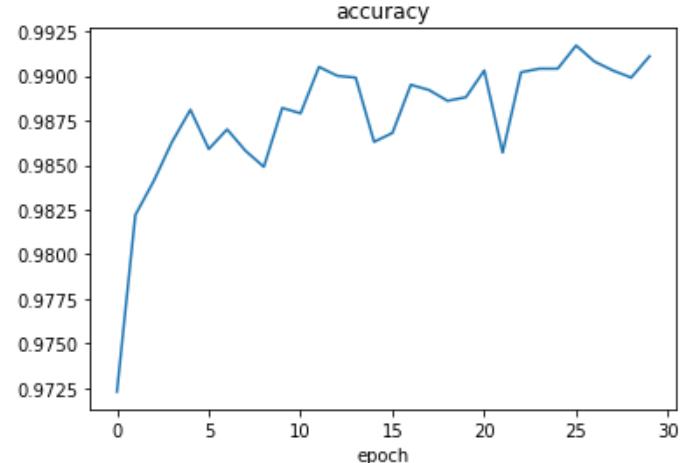
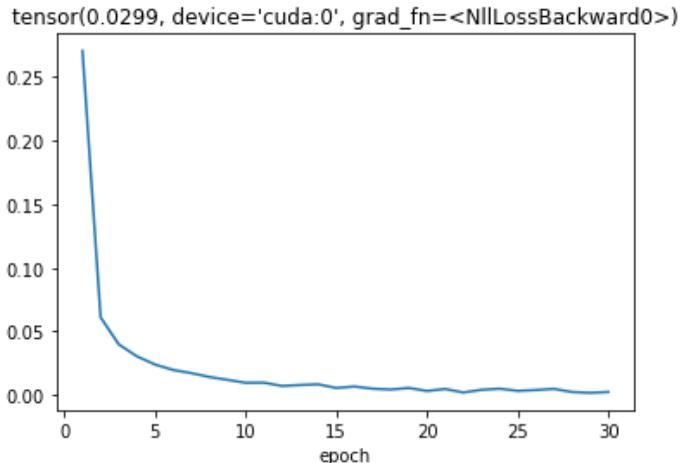
Model 1. 2 Conv, 1 Max pooling, 2 FC

좌: train loss 우: test accuracy



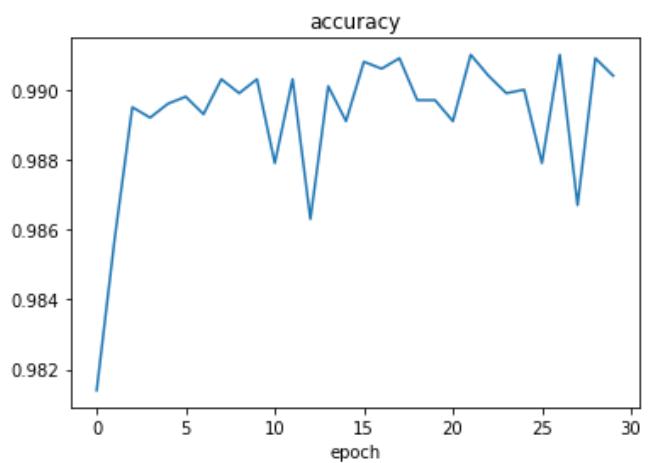
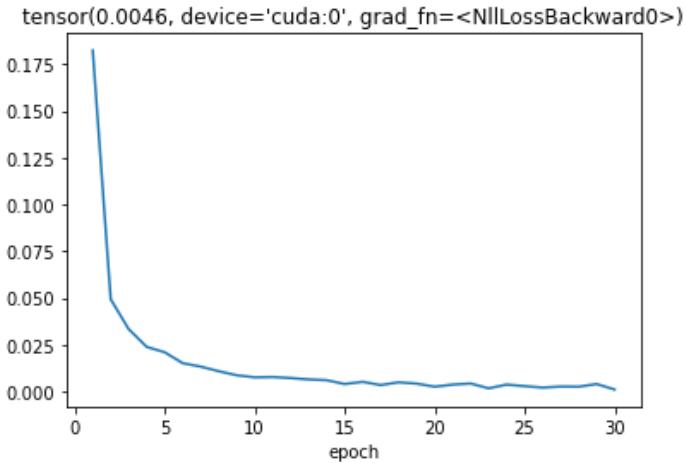
Model 2. 4 Conv, 2 Max pooling, 2 FC

좌: train loss 우: test accuracy



Model 3. 5 Conv, 1 Max pooling, 2 FC

좌: train loss 우: test accuracy



	Final training loss	Final test accuracy
Model 1	0.0025606804678304364	99.01%
Model 2	0.0038296734294880016	99.24%
Model 3	0.0010124977973414021	99.04%

- Training loss를 average loss of the number of batch로 설정함. (batch size=100, batch 개수는 600)

분석

Parameter 개수 비교. (Model 1 vs Model 2)

Model 1	#parm	Model 2	#parm	Model 3	#parm
Conv 1	1*3*3*32	Conv 1	1*3*3*8	Conv 1	1*3*3*32
Conv 2	32*3*3*64	Conv 2	8*3*3*16	R.B_Conv 2	32*3*3*32
FC 1	61*14*14*64	Conv 3	16*3*3*32	R.B_Conv 3	32*3*3*32
FC 2	64*10	Conv 4	32*3*3*64	R.B_Conv 4	32*3*3*32
		Fc 1	64*7*7*64	R.B_Conv 5	32*3*3*32
		Fc 2	64*10	Fc 1	32*14*14*64
				Fc 2	64*10
total	819,176	total	225,608	total	439,200

필요 연산량 Model1 > Model3 > Model2

세 모델 중 누가 가장 좋은 퍼포먼스를 보여줄 것인가에 대한 대답은 모델2와 모델3이 학습 조건 하에서 달라진다고 대답할 수 있다. 하지만 어느 모델이 가장 퍼포먼스가 좋지 않느냐에 대한 대답은 확실히 할 수 있다. 모델1이 그 대상이다.

모델1의 경우 #parm이 모델2에 비해서는 약 4배 모델3에 비해서는 약 2배에 가깝다. 연산량이 많다고 해서 모델의 성능이 우수하지도 않다. Accuracy가 98퍼센트에 도달하는 epoch 횟수가 다른 두 모델에 비해 좀 더 길다. Train loss 그래프를 보면 기울기가 확연하게 차이나는 knee부분은 [2,5] 사이이다. 다른 두 모델은 epoch [2,5] 부근에서 98.후반%를 보여주고 있는 반면에 모델1은 최대 98.초중반%를 보여주고 있다. 연산량도 많아 학습 비용이 많은데 성능은 두 모델에 못미친다는 증거다.

나머지 두 모델에 대해 얘기하자면 모델 2의 경우에는 epoch [0,10] 구간에서는 accuracy가 모델3보다 낮은 걸 확인할 수 있다. Final accuracy는 모델2가 모델3보다 근소하게 높다.

즉 낮은 epoch에서는 ResNet구조를 채용한 모델3가 더 좋다고 생각한다. Epoch 20 이후 구간에서는 모델2가 근소하게 낮다고 생각한다.

만약 30 epoch가 아닌 더 큰 epoch를 사용한다면 Gradient vanishing을 방지할 수 있는 ResNet구조를 사용한

model3을 사용하는 게 가장 좋다고 생각한다.