Self-supervised Learning for image classification on DeepFashion

Dusan Mihajlov, Pinar Goktepe, Michael Baur

May 21, 2019

Plan

1) Dataset
2) Preprocessing
3) Self Supervised Learning
4) Classification
5) Results
6) Conclusion

1.)Dataset: DeepFashion

- 289222 images
- 48 classes

2) Preprocessing

- Actual Image size: up to 300x300
- Random crop to the smaller size (width or height)
- Resize to 227x227
- Generate a random permutation of patches
- Make a new image with patches ordered in the same way as permutation
- Declare labels as 2D coordinates of each patch
- Number of outputs = number of patches * 2

3) Self-supervised Learning

- Dataset: 70% of the initial dataset
- Task: Jigsaw Puzzle (2x2 puzzle or 3x3 puzzle)
- Model: AlexNet

- Input: ---> Output: [(1,0),(0,2),(1,2),(1,1),(0,1),(2,2),(0,0),(2,0),(2,1)]

Parameters: Optimizer=Adam, lr=0.0005, epochs=20, scheduler=StepLR, step size = 5, gamma=0.3, weight decay=0.001, tile size=75 or 113 (75 for 3x3 puzzle and 113 for 2x2 puzzle), loss function=MSE

4) Classification

- Dataset: 30% of the initial dataset
- Task: Clothing Classification
- Model: AlexNet

- Input :

-----> Output: category #5

4) Classification

- After Solving Pretext Task
  - AlexNet initialized with the weights of the self-supervised part
- With ImageNet Pretraining
  - AlexNet initialized with the weights of the ImageNet pretraining
- Without Pretraining and Pretext Task
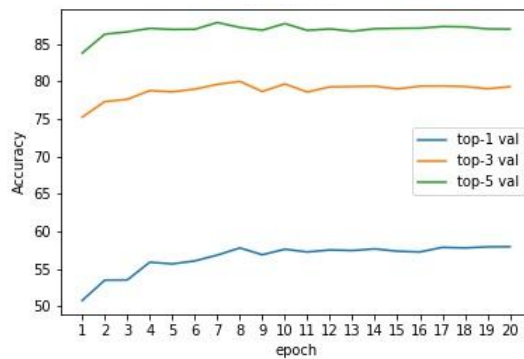 - AlexNet initialized with random weights

Parameters: Optimizer=Adam, lr=0.0001, epochs=20, scheduler=StepLR, step size = 5, gamma=0.3, weight decay=0.001, loss function=CrossEntropyLoss
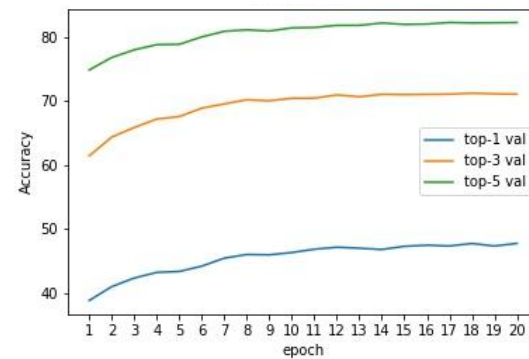
5) Results: Validation Accuracy Curves

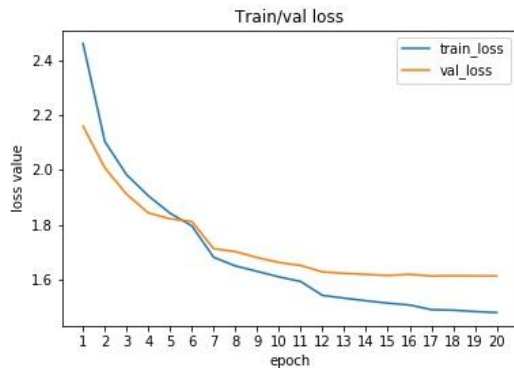Without Pretraining and Pretext Task Solving
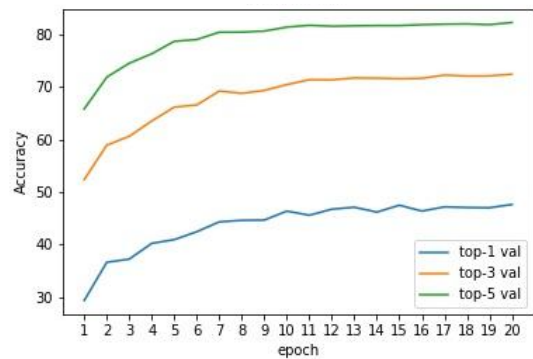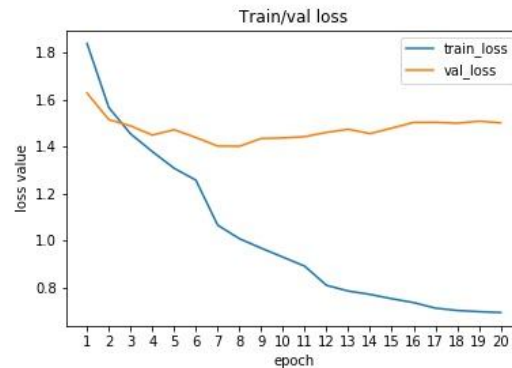
With Pretraining

With Pretext Task Solving
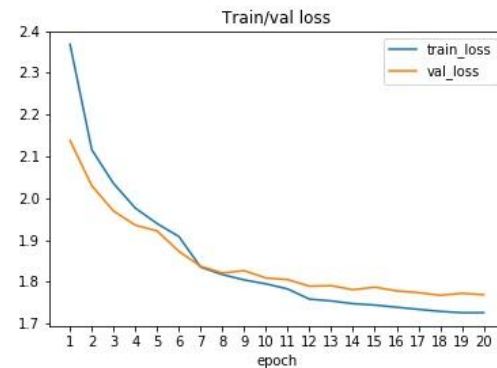
5) Results: Training and Validation Loss Curves

Without Pretraining and Pretext Task Solving

With Pretraining

With Pretext Task Solving

5) Results: Testing Results

Test Accuracies

Model	Top-1 Test Accuracy	Top-3 Test Accuracy	Top-5 Test Accuracy
Classification with Alexnet	52.175	75.467	85.173
Pretext Task+Classification with Alexnet	47.364	71.469	82.164
Classification with Pretrained Alexnet	59.459	81.184	88.593

6) Conclusion

Possible Drawbacks of Self Supervised Learning:

- Different focuses of pretext task and target task
- Need for deeper network architectures
- Dataset size
- Longer training procedure for pretext task

Questions?