CurriculumNet: Weakly Supervised Learning from Large-Scale Web Images

Jeffery Zhang

October 29, 2018

Problem

Deep Neural Networks can achieve great performance, but require large amounts of accurately labelled data. Obtaining accurately labelled data is very labor intensive and very expensive.

Problem

We would like to be able to obtain labelled data for supervised learning in a cheaper manner. One potential source of data is the web. For example: the images returned via web queries can be thought of as being annotated by the query, this is a form of weak supervision.

Problem

The problem with this approach is that web query results are noisy. Noisy labels degrade machine learning performance. We need an approach to learning that can deal with noisy data.

Solution

We propose an approach that improve the learning capability of standard neural networks by using a novel training strategy: Curriculum Learning.

Specifically, we study the problem of learning CNN from large scale image data sets with large amounts of noisy labels. The goal is to be able to handle noisy labels and data imbalance effectively.

Solution

Step 1: Initial feature generation

Use all training data to learn an initial model (Inception v2) which is then applied to computing a deep representation from each image in the training set.

Solution

Step 2: Curriculum Design

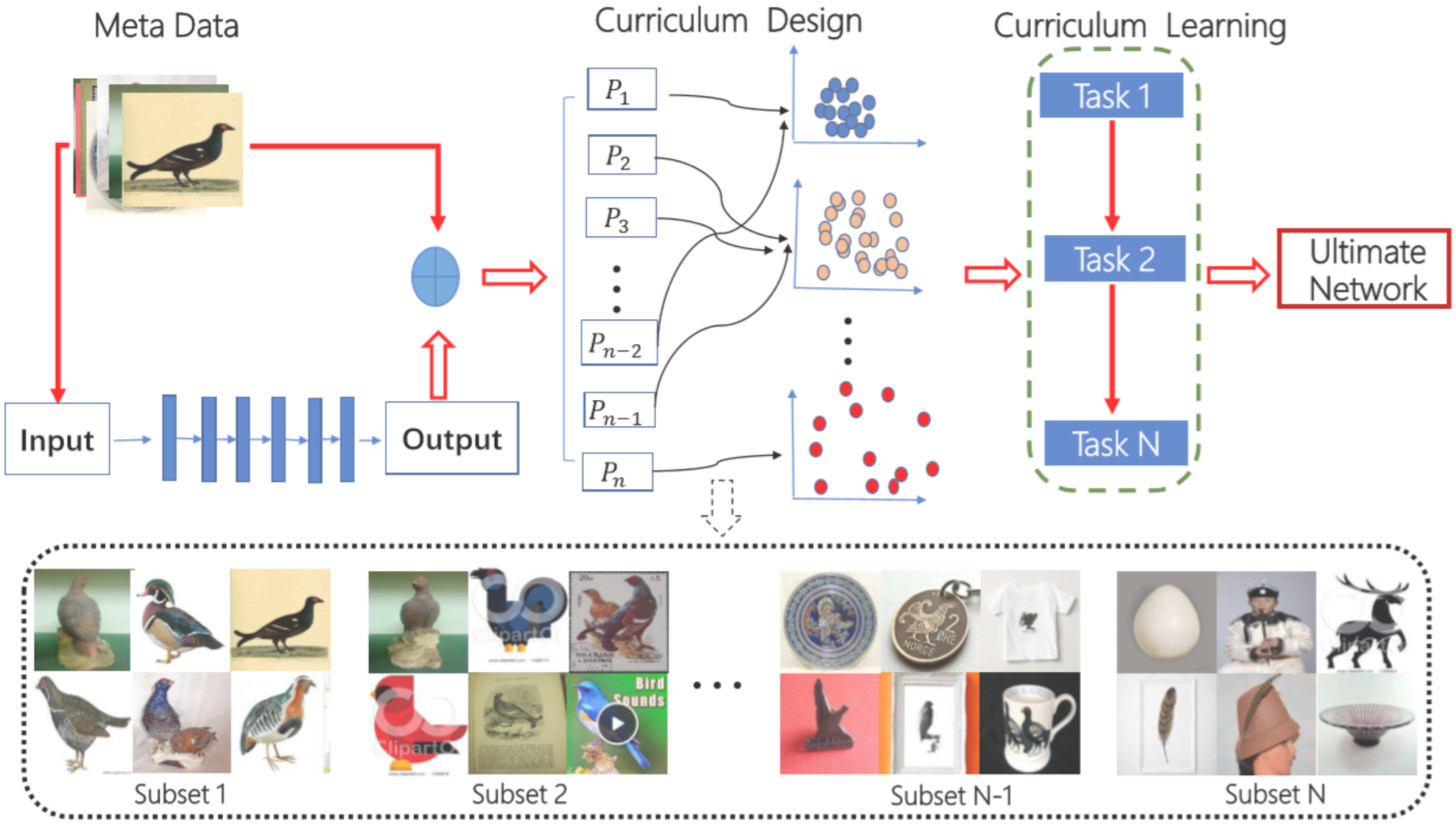
Use the initial model to map all training images into a feature space where the underlying structure and relationship of the images in each category can be discovered

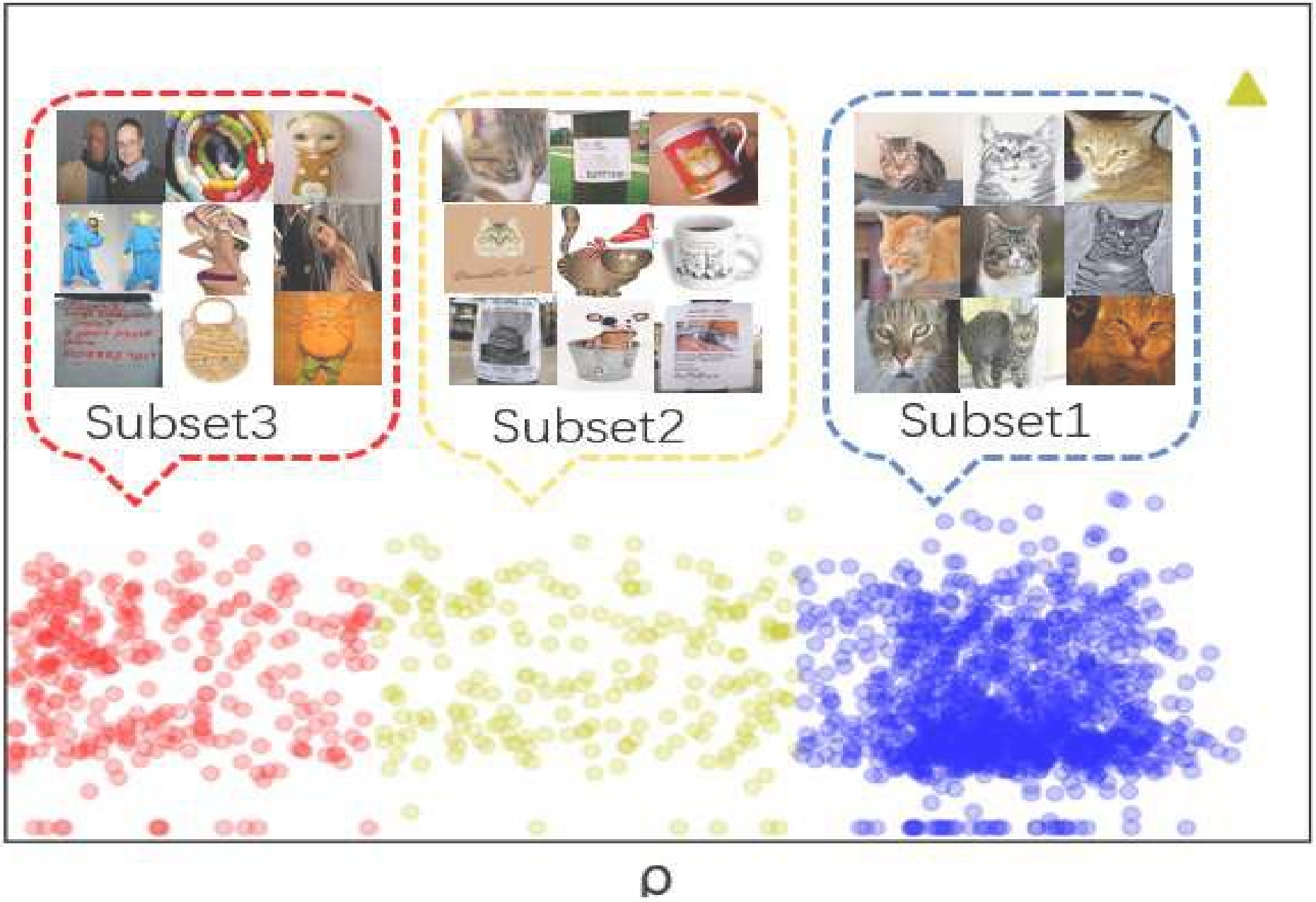
We define a complexity metric over the representation space, and cluster the examples into buckets by complexity

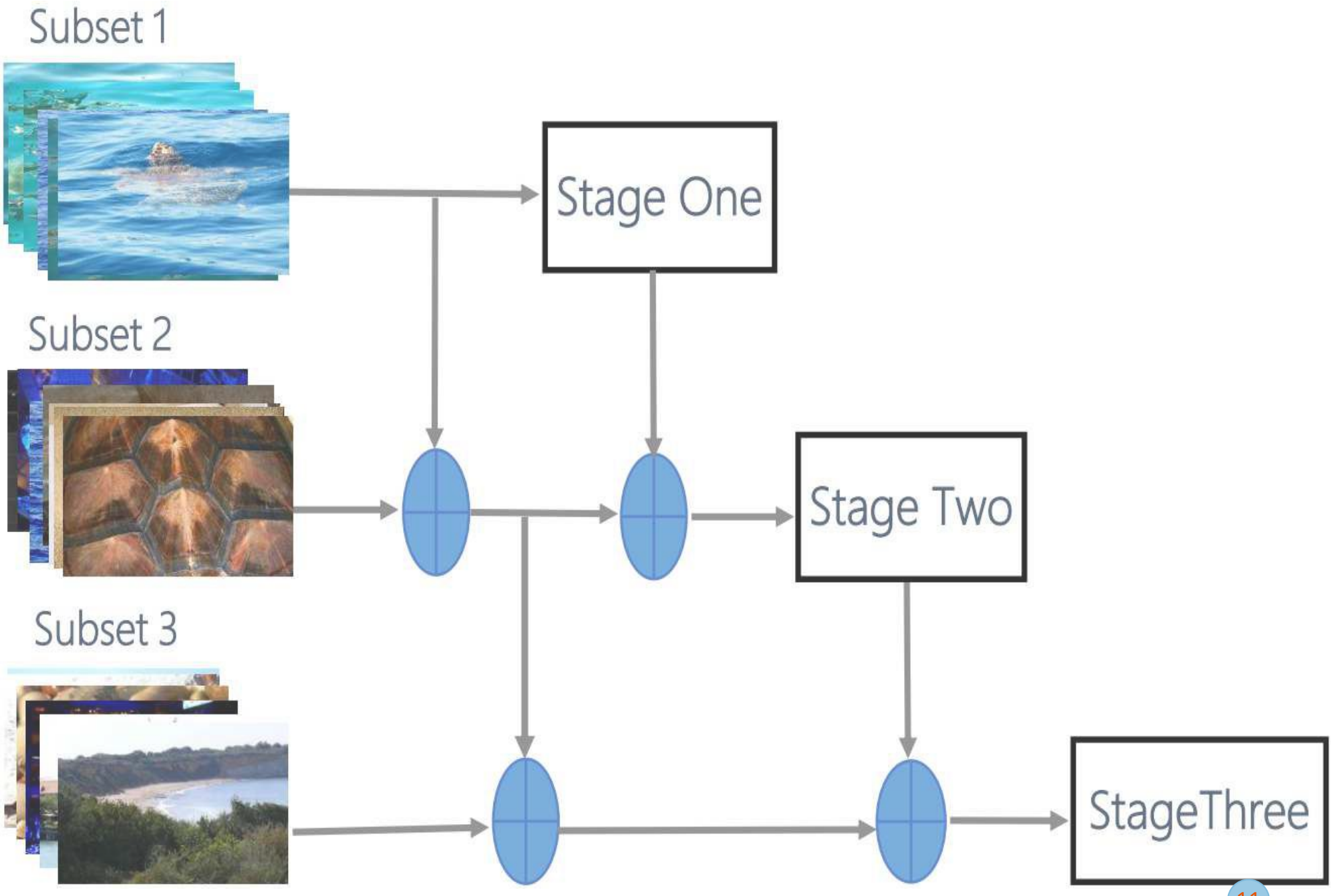
Solution

Step 3: Curriculum Learning

Train the CNN starting with the easiest bucket of examples and progressively adding examples from more complex buckets


Meta Data
Curriculum Design
Curriculum Learning
P_1
P_2
P_3
P_{n-2}
P_{n-1}
P_n
Task 1
Task 2
Task N
Ultimate Network
Input
Output
Subset 1
Subset 2
Subset N-1
Subset N

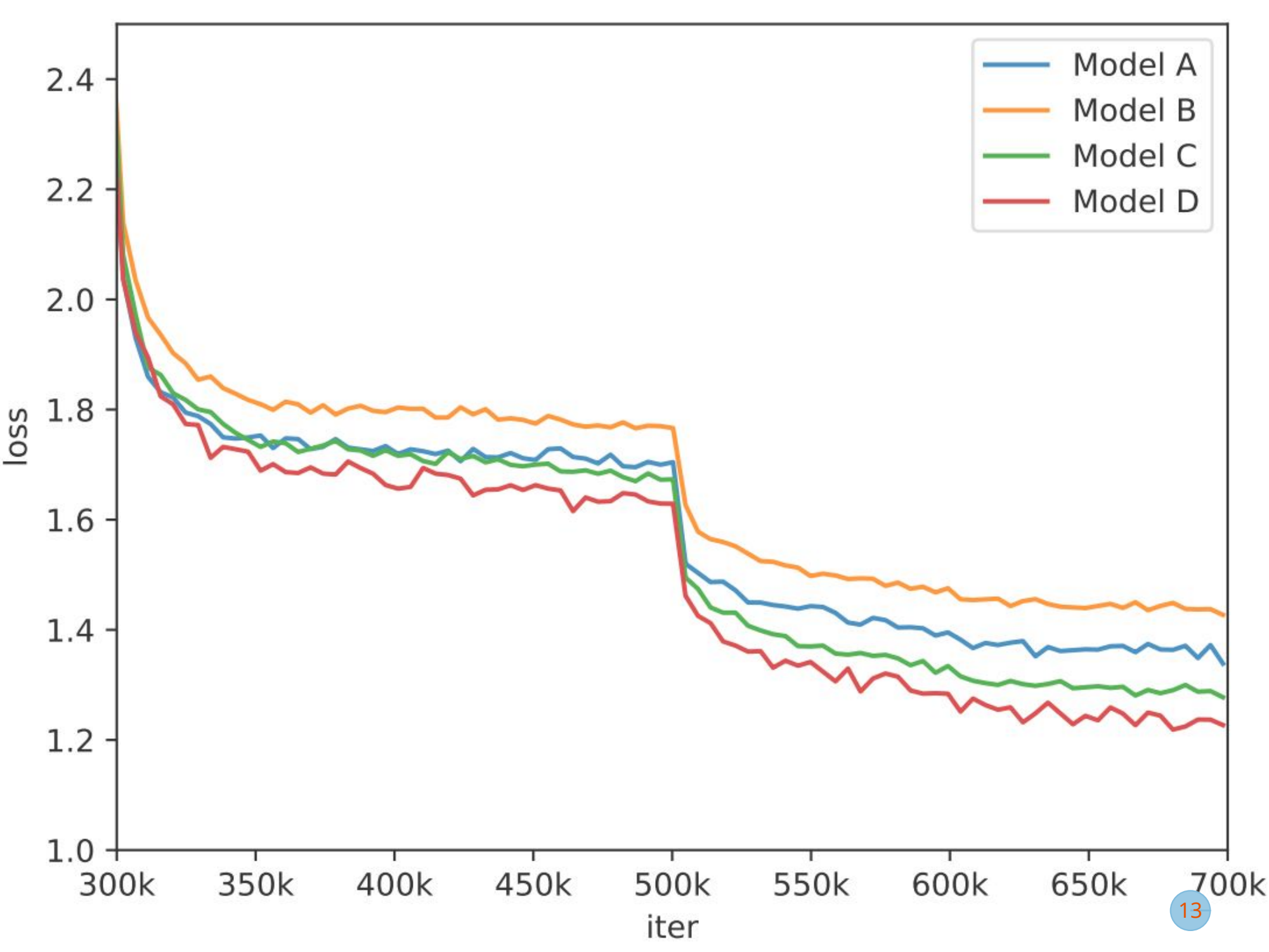

Subset3
Subset2
Subset1


Subset 1
Stage One
Subset 2
Stage Two
Subset 3
StageThree

Results

We evaluate 4 different training strategies using a standard Inception v2 architecture

Model-A: trained directly using whole training set
Model-B: trained only using clean subset
Model-C: trained using proposed strategy with 2-subset curriculum: clean and noisy subsets
Model-D: trained using proposed strategy with 3-subset curriculum: clean, noisy and highly noisy subsets


Model A
Model B
Model C
Model D
loss
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
300k
350k
400k
450k
500k
550k
600k
650k
700k
iter

Questions

Questions?