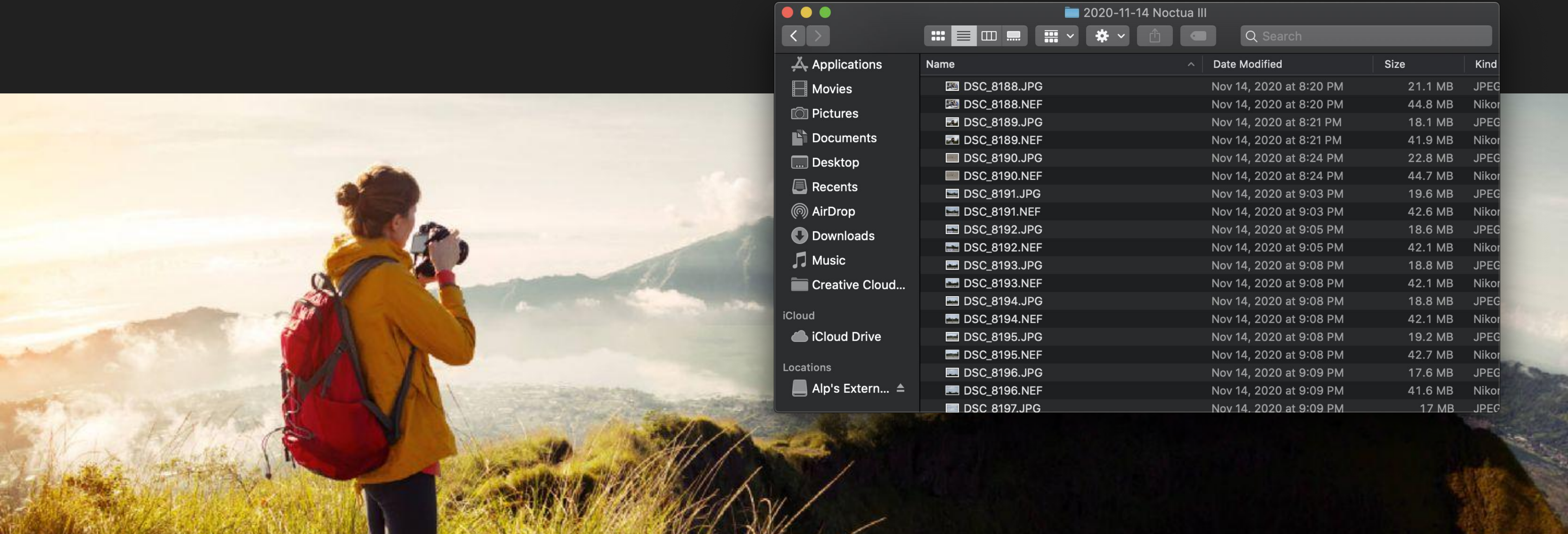
Landmark Recognition

A multi-class image classification problem.

Problem Statement

Are you a travel photographer who’s photo archive look like this? Thousands of rows of DSC_XXXX.jpg files...



Problem Statement

Landmark recognition can help label the photos you take while travelling, making your travel archives more human-readable. We aim to predict labels through image pixels.

Thus, you can understand your archives and search for the picture you have taken years ago easily.

Team WAM (Will-Alp-Mason)

Will Ledig

COMP, DSCI

Hanszen ‘21

- Preprocessing
- Preliminary Data Analysis
- PyTorch Implementation

Mason Reece

SOPA, POLI, DSCI

Hanszen ‘22

- Preliminary Data Analysis
- Literature Search & Algorithms
- PyTorch Implementation

B. Alp Yakici

COMP, DSCI

Hanszen ‘22

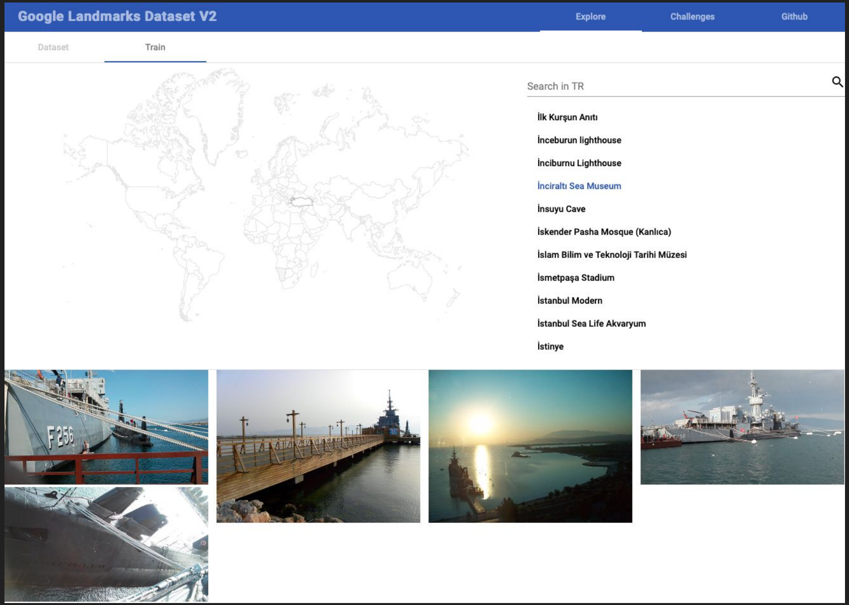
- HPC Person
- Presentation
- Literature Search & Algorithms

Dataset

We use Google Landmarks Dataset v2 (GLDv2) for our training set.

Advantages

- Publicly available.
- Easy to download (comes with a Shell script that downloads from AWS S3).
- Comes with online dataset explorer.
- 98.27 GB worth of labeled data.
  - Corresponds to 3.2 million images.



Dataset explorer on Google APIs.

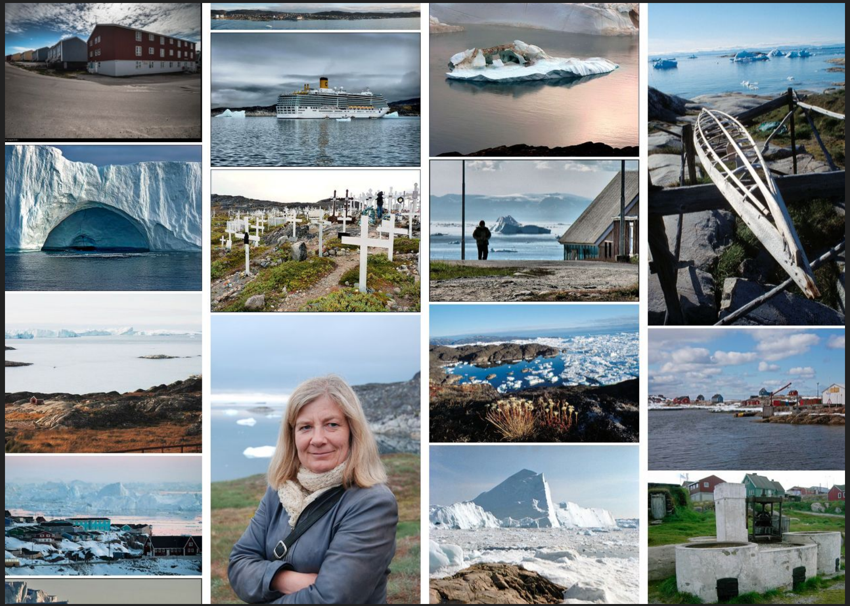
"Google Landmarks Dataset v2 - A Large-Scale Benchmark for Instance-Level Recognition and Retrieval"
T. Weyand*, A. Araujo*, B. Cao, J. Sim
Proc. CVPR'20

Dataset

We use Google Landmarks Dataset v2 (GLDv2) for our training set.

Disadvantages

- Very unwieldy.
- Some labels are very generic.
    - Several landmarks have wide range of photos with different subjects.
- Wide range of dimensions.
    - Need for a significant preprocessing pipeline.



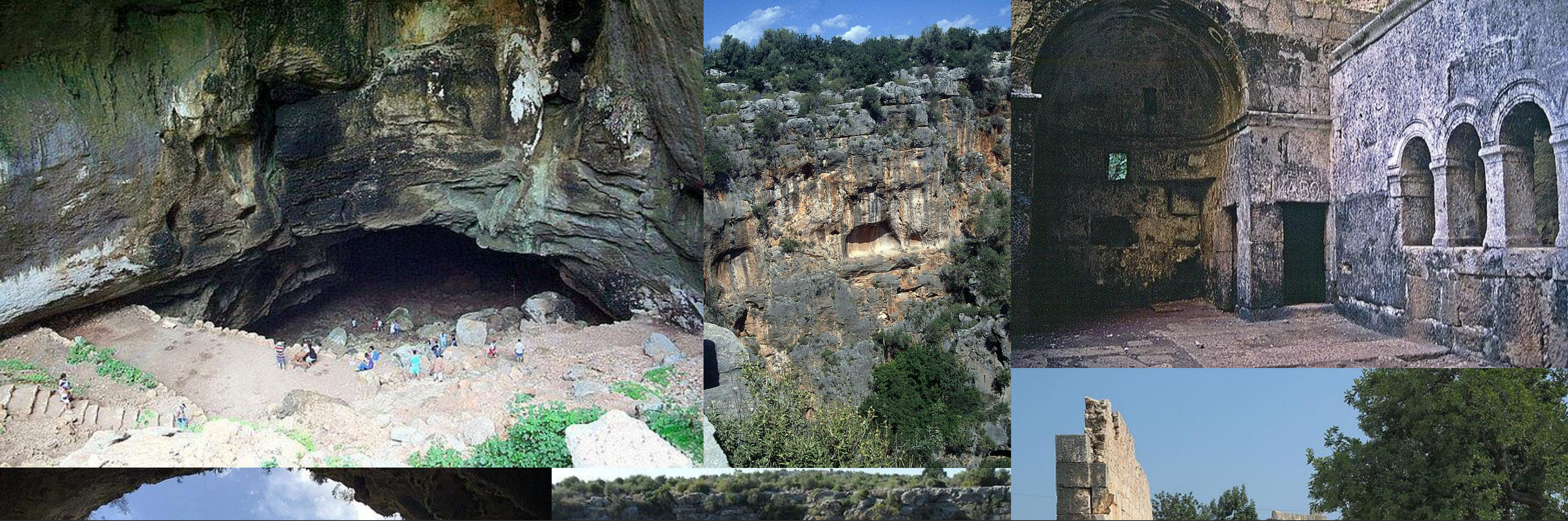
Images that are classified as “Illulissat,” which is a municipality in Greenland.

"Google Landmarks Dataset v2 - A Large-Scale Benchmark for Instance-Level Recognition and Retrieval"
T. Weyand*, A. Araujo*, B. Cao, J. Sim
Proc. CVPR'20



Wide Range of Labels / Dimensions

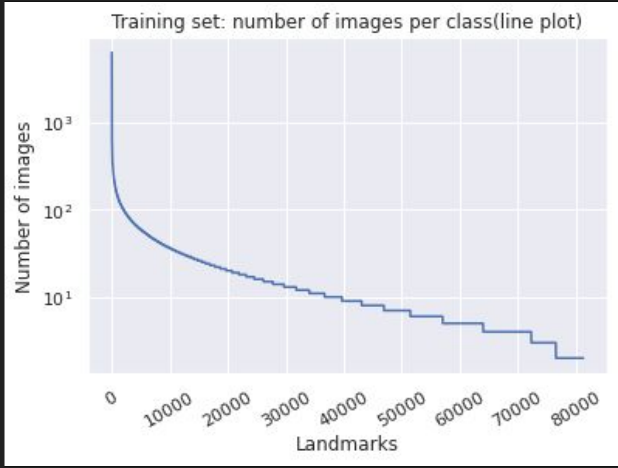


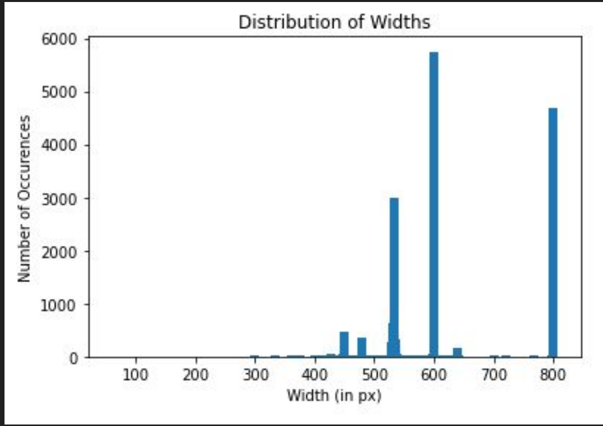


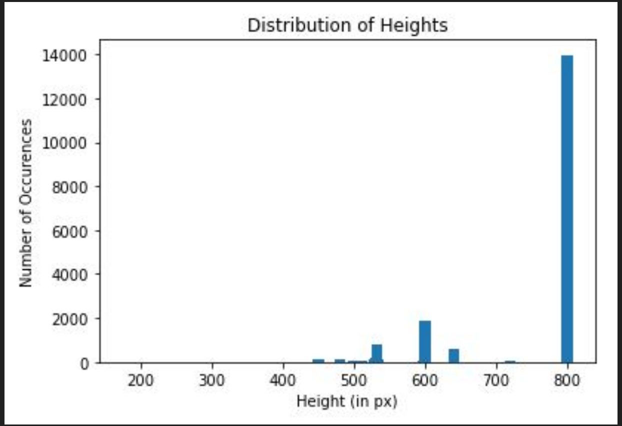
Wide Range of Data for One Label

Preliminary Data Analysis

- Only about half of the landmarks (40,000) had more than 10 images in the dataset.
- Less than 2000 landmarks had more than 100 images.
- Majority of images with larger height than width.
- Most common dimensions were 800 x 600 px.







Preprocessing

- Decided to only use landmarks with at least a set number of images (i.e. 250+, 500+, 1000+).
- Split our training data into train and validation set, 80-20% split.
- Resizing to 48x48 pixels, random horizontal flipping.
- Considered more transformations, but due to already noisy, minimal data, we decided not to do so.

Models & Algorithms

Models & Algorithms

Deep Convolutional Neural Networks (DCNNs)

- **Pretrained Feature Extractor:** EfficientNet B7
- **Loss Function:** ArcFace
- **Optimizer:** Stochastic Gradient Descent (SGD)
- **Adjusting Learning Rate:** Cosine Annealing
- **Verification/Evaluation:** Micro Average Precision (μAP)

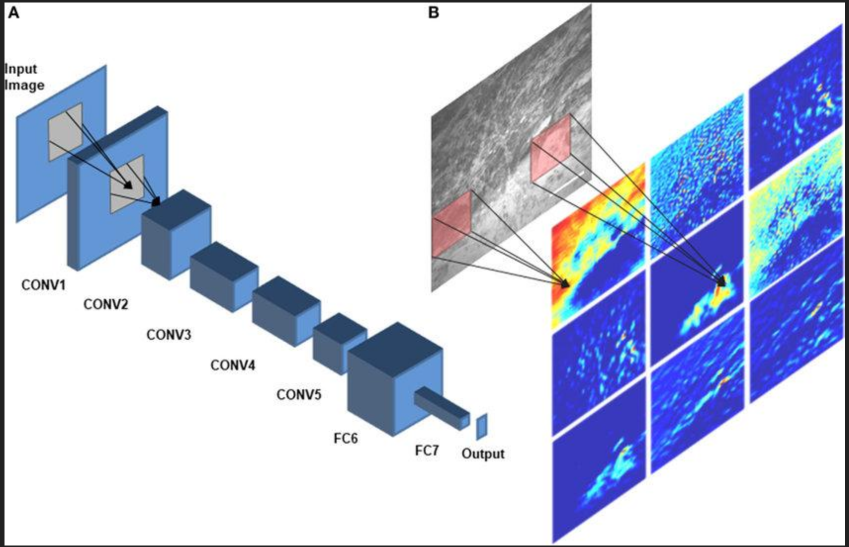
Deep Convolutional Neural Networks (DCNNs)

Pre-Trained Feature Extractor: EfficientNet by Google

- A family of image classification pre-trained models.
- Originally for Tensorflow, however:
  - PyTorch implementation is provided by effnet-pytorch module.
- Used B7.

Advantages

- Publicly available on GitHub.
- Pre-trained, relatively small, and efficient.



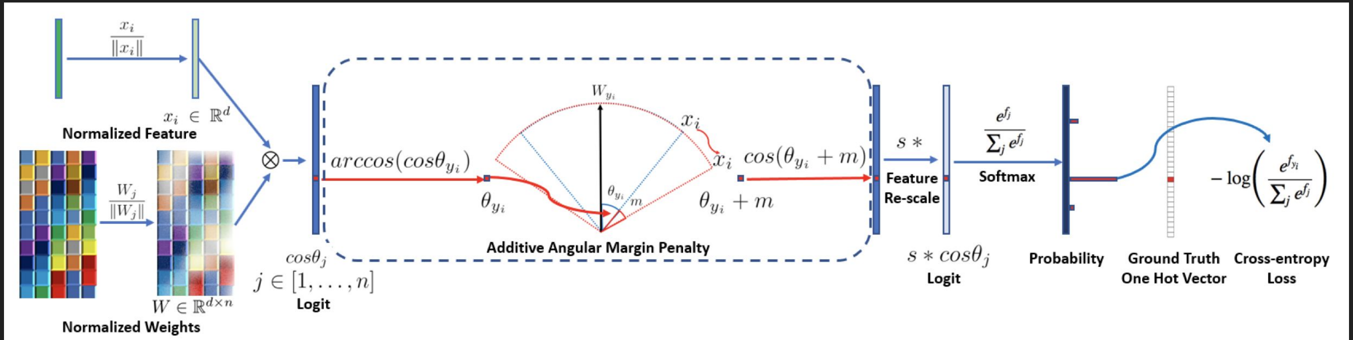
DCNN Architecture

M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Proc. CoRR'19

Loss Function: ArcFace

- ArcFace: Additive Angular Margin Loss for Deep Face Recognition
- Improves the discriminative power of the model.
    - Makes it easier to distinguish different landmarks.

- Code is published online on GitHub.



J. Deng, J. Guo, N. Xue and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," 2019

Optimizer: Stochastic Gradient Descent (SGD)

- Performs parameter update for each training image and its associated label.

Advantages

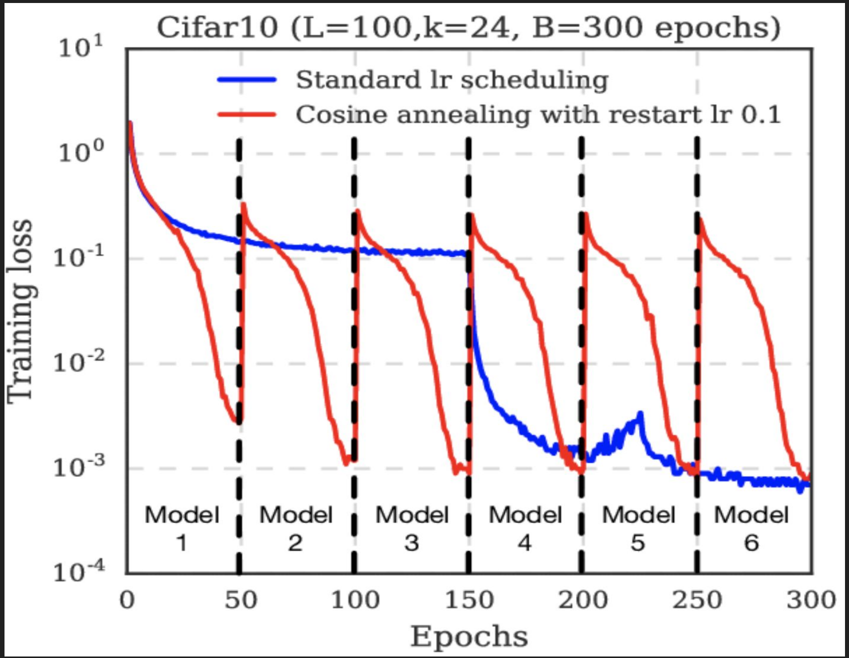
- Easy to implement and fast.
- Doesn’t require a huge amount of memory.
- Learns online.
    - Allows us to compute best weights at each epoch.

Disadvantages

- High variance, results in our function to fluctuate.
- Requires more epochs.
    - Hence, we run the training on HPC for about 9:00:00 hrs.
    - We were not able to do that yet.

+ **Adjusting Learning Rate:** Cosine Annealing

- Start with a large learning rate (LR).
- Rapidly decrease LR to a minimum value before rapidly increase again.
- Resetting of the LR acts like a *simulated* restart of the learning process.
- This is called a “warm restart,” as we reuse good weights as the starting point of the restart (in contrast to “cold restart”).



I. Loshchilov and F. Hutter, "SGDR: STOCHASTIC GRADIENT DESCENT WITH WARM RESTARTS," Proc. ICLR'17

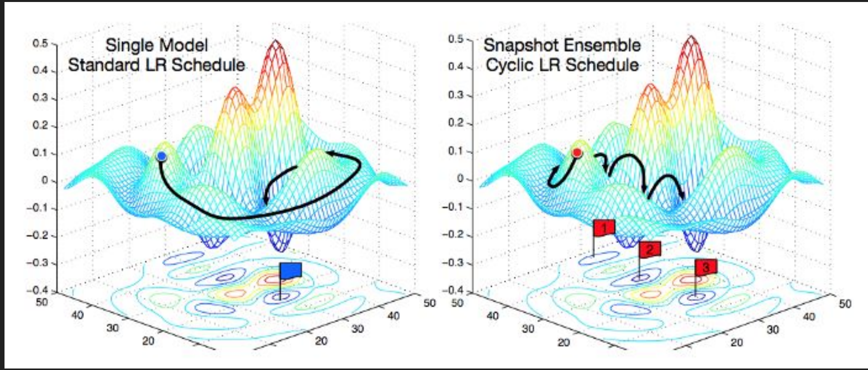
+ Adjusting Learning Rate: Cosine Annealing

Advantages

- Increases the chances of finding a much better minimum.
  - “Jumps” out of a local minimum that is far from the global minimum.

Disadvantages

- Requires relatively high number of epochs to outperform standard LR schedules.



I. Loshchilov and F. Hutter, "SGDR: STOCHASTIC GRADIENT DESCENT WITH WARM RESTARTS," Proc. ICLR'17

Evaluation

Micro Average Precision (µAP) is used for evaluating our model. µAP measures both ranking and the consistency of the scores across multiple queries.

- Foreach test image, predict one label and a confidence score.
- Sort the list of predictions in descending order by the confidence scores.

$$\mu AP = \frac{1}{M}\sum_{i=1}^{N} P(i)\, rel(i)$$

F. Perronnin, Y. Liu, and J.-M. Renders, "A Family of Contextual Measures of Similarity between Distributions with Application to Image Retrieval," Proc. CVPR'09

Micro Average Precision (µAP)

$$\mu AP = \frac{1}{M} \sum_{i=1}^{N} P(i)\, rel(i)$$

Parameters:

- M: Total number of queries with >= 1 landmark in the training set.
- N: Total number of predictions across all queries.
- *P(i)*: Precision at rank i.
- *rel(i)*: Relevance in prediction i.
  - It is 1 if the i^{th} prediction is correct. Otherwise, it is 0.

F. Perronnin, Y. Liu, and J.-M. Renders, "A Family of Contextual Measures of Similarity between Distributions with Application to Image Retrieval," Proc. CVPR'09

Micro Average Precision (μAP)

Why μAP?

- μAP considers consistency of the scores.
  - Thus, using Micro Average Precision is preferable since there are label imbalances, as we observed in our preliminary data analysis.
- Ranking can be computed asynchronously with accumulators.
  - Evaluation is performed faster.

F. Perronnin, Y. Liu, and J.-M. Renders, "A Family of Contextual Measures of Similarity between Distributions with Application to Image Retrieval," Proc. CVPR'09

Results

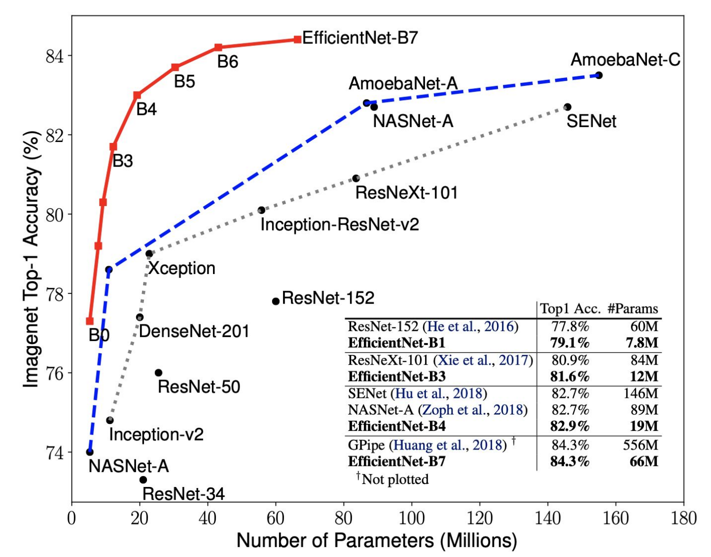
+ Alternative Models & Algorithms

µAP Score

- Trained small subset of the dataset on our local desktop.
  - 48 x 48px Images
  - 1 x Nvidia GTX 2060
  - 7 Epochs
  - Result: µAP = ~0.15 - 0.20.
- We expect µAP = ~0.40 - 0.50.
  - On 128 x 128px Images
  - Rice NOTS HPC Clusters (2 x Tesla V100 or Tesla K80)
- Best solutions achieve µAP = 0.60 - 0.65.

Pretrained Model Options

- VGG-16
  - 74.5% top-1 test accuracy in ImageNet dataset.
  - Large model size.
- Residual Network (ResNet)
  - Winner of 2015 ImageNet challenge.
  - Deeper than VGG-16 and smaller size.
- EfficientNet
  - 84.3% top-1 test accuracy in ImageNet dataset.
  - Even smaller.



Lack of time to test ourselves, relied on prior work, smaller model size, and ease of implementation lead us to choose EffNet B7.

Loss Function Selection

Loss	LFW	CFP-FP	AgeDB-30
Softmax	99.7	91.4	95.56
SphereFace (m=4, $\lambda = 5$)	99.76	93.7	97.56
CosineFace (m=0.35)	99.80	94.4	97.91
ArcFace(m=0.4)	99.80	**94.5**	98.0
ArcFace(m=0.5)	**99.83**	94.04	**98.08**

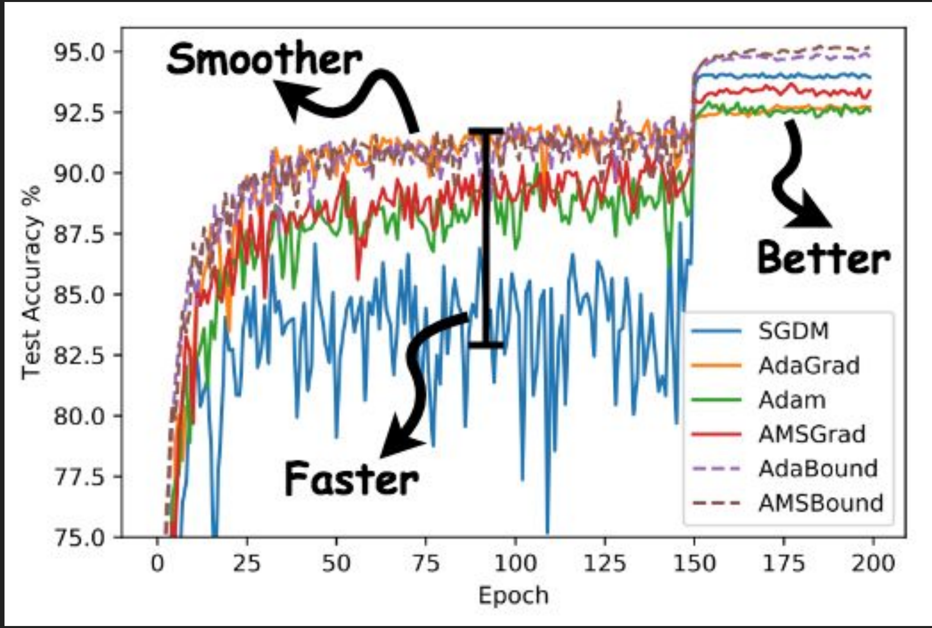
Verification performance (%) of models on the following datasets:

- LFW: Labeled Faces in the Wild
- CFP-FP: Celebrities in Frontal-Profile in the Wild
- AgeDB-30: In-the-Wild Age Database

J. Deng, J. Guo, N. Xue and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," 2019

Optimizer Selection

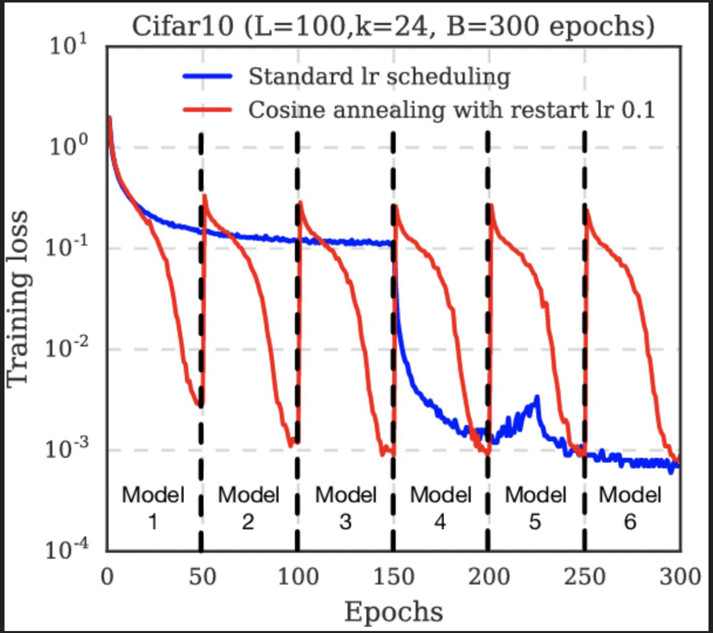
- Adam
  - Smoother.
- SGD
  - Faster.
- AdaBound
  - Mix.



D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2017

Adaptive Learning Rate Selection

- Static LR
  - Easy to implement.
  - Not optimal.
- Jittering
- Step-LR
  - To decrease slowly.



I. Loshchilov and F. Hutter, "SGDR: STOCHASTIC GRADIENT DESCENT WITH WARM RESTARTS," Proc. ICLR'17

Conclusion

Improvements & Future Work

- Larger images (128 or 256 px).
- Multiple model comparison and ensemble methods.
- Training on larger epochs.
- Schedule Multi-GPU NOTS jobs.

berkalpyakici — bay1@nlogin2:~ — ssh bay1@nots.rice.edu — 100×24

```
[bay1@nlogin2 ~]$ squeue -u bay1
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
           2142710   commons DSCI303L     bay1 PD       0:00      1 (QOSGrpCpuLimit)
[bay1@nlogin2 ~]$
```

Hopefully, our model will be trained on NOTS one day...

Thank You