Funding Proposal for AI-Face

MATH 597-010 Semester Project

Team: **Sahil Mishra** (szm5707@psu.edu), **Tom Kawchak** (tyk5178@psu.edu) & **Olga Dorabiala** (omd5024@psu.edu)

Summary

We are requesting $800 to use for using Google's cloud services to build and train our model for the Megaface Challenge. This includes estimated costs ($711.88) and a bit extra in case there are unforseen costs or we use the machine more than expected. We estimate that training for a week should be enough to give us sufficient results to evaluate the effectiveness of the model.

Project Description

Face verification is perceived, by many, as a solved problem. However, when the state-of-the-art models are tested at the million-scale they perform with dramatically varying accuracies. In this work we evaluate the performance of the Range Loss Function with Residual Networks for face verification on the MegaFace Dataset.

To address the questions of varying accuracies of the model when tested on a bigger dataset we test the state-of-the-art models on the MF2 dataset, that requires all algorithms to be trained on same data, and tested at the million scale

MF2 is a large-scale, publicly available dataset with **672,057** unique identities and **4,753,520** photos. The minimum photos per identity is 3 while the maximum is 2,469. The mean photos per identity is **7**.

Requirements & Justification

We would like more funding to continue to use Google Cloud to train our model on the MegaFace Dataset. Due to the size of the dataset, we are unable to load the data onto our computers, and therefore cannot run training. So far, we have been using the free trial that Google offers to students, but it has run out. Therefore, we are requesting more funding, since we already have the data in the Google Cloud Instance, and we would like to minimize any need to change machines / services.

Pricing and Project Usage

Service	Cost	Usage Prediction	Total Cost ($)
Storage	0.0170 ($ per GB-month)	1TB SSD for 1 month	170
Machine*	1,864.63 ($ per month)	1 weeks for training	466.16
Snapshot of data	0.002572 ($ per GB-month)	1TB	25.72
Miscellaneous**			50
Totals			711.88

*Machine:

- 12 CPUs with a total of 78 GB memory
- 4 NVIDIA Tesla K80 GPUs

**Miscellaneous: Various other overhead that Google charges for such as:

- Compute Engine Network Internet Egress from Americas to Americas
- Compute Engine Custom instance Ram running in Americas

We Need at Least 75 GB of memory because the model itself is 75GB of memory space. 12 CPUs will tremendously help to speed up the computation time, along with the 4 NVIDIA GPUs. We found 4 GPUs to be a reasonable balance between cost and performance so that we still train the model in a reasonable time period. Storage requirements of 1TB are based upon the dataset size (about 900GB).

Payment Information

For Penn State Account
name: Thomas Kawchak
PSU alias: tyk5178
PSU ID: 944339690

email: tyk5178@psu.edu
phone: (610) 312-1220

Reach out over email (tyk5178@psu.edu) or phone (610-312-1220) if you have any questions, need me to come by your office, or need any additional information.