Image Recognition Training via Reddit

Computer Vision Project Work Survey

Kenneth Sanders

May 3, 2018

Introduction

Social media and link aggregation have become critical tools for targeted data acquisiton, which in turn is important for training modern "AI" networks. This project, Image Recognition Training via Image Sharing Websites, aims to develop a novel and simple tool for scraping labeled images from Reddit, saving them, and feeding them into any number of image processing networks. This tool is aimed at personal and small projects in order to abide my Reddit's and Imgur's (the main filehost for Reddit) TOS (Terms of Service). Unfortunately, the idea of using both social-media sites and crowdsourcing to generate image recognition datasets is not new in and of itself. Many others have had the same idea, with varying way of implementing it. From the facial recogntion using user tags from Facebook, to the Captcha methodology of generating training sets. Additionally, the design and implementation of the Convolutional Neural Networks is generally dependent on the quality, consistency, and size of the training dataset, and any dataset generator should take into account for a need for a large range of those dataset properties. Therefor, this paper takes the opportunity to detail some of the more influential work by others that we have built upon.

Previous Works

The following sub-sections detail the content of the works from which our project draws information as well as what information in was used in particular.

Multi-column deep neural networks for image classification

by J. Schmidhuber, U. Meier and D. Ciresan

This report details the structure and training of a complex convolutional neural network based on biological models as well as methods to evaluate its perfomance. Their model is based around small, winner-take-all trained convolutional layers. They also show better-than-human results on MNIST handwritten digit recognition, and twice as accurate results for traffic sign identification. Our project aims to use some of the notes and applications their models in order to better structure the our network models to fit the experiment. In particular, the small convolutions and relatively large networks depth is a great example network to test witht the data acquisition methods in our project,since it is readily trained on a small number of categories, which is paired well with our data collection [SMC12].

Rethinking the Inception Architecture for Computer Vision

by Christian Szegedy et al.

In this paper, the authors discuss the efficiency of traditional CNNs, and how increasing the size of input training samples and the respective network to process that sample affects the computation time for the network. This is extremely important to my project, since the data provided by Reddit will generally be high resolution, so the choice of how much to downscale makes a large difference int raining times for each of the evaluation CNNs. This paper then goes on to discuss the efficiency of the previous iteration of the Inception Architecture, and details how the third iteration of their CNN is modeled. Lastly, they describe their method of testing their latest iteration on the ILSVRC12 classification challenge, and show how they obtained a 5-20% increase in efficiency over the state of the art at the time [Sze+15].

Image Labeling on a Network: Using Social-Network Metadata for Image Classification

by Julian McAuley and Jure Leskovec

This paper describes how using social media metadata as further inputs for image classification via structured learning. The paper takes some image recogntion dataset that were gathered on Flickr and attempts to apply various metadata properties to the datasets to add further relational features to it. They predict that this will have positive effects on the model training to improve image classification beyond standard image-only models. Some of these metadata properties may also be found on Reddit, such as: the comment thread, who uploaded the image, from where, and the friends of the original poster. These properties could be used to further expand our model to gather metadata as well as the labeled images, which may give more context and information about the images and the categories they belong to [ML12].

Learning by expansion: Exploiting social media for image classification with few training examples

by Sheng-Yuan Wang et al.

This paper is very close to the topic of our project. It displays great accuracy for crowd sourced training images in image recognition details both pros and cons to this methodology. They show how the use of additional contextual (meta-)data may be used to aid in semantic graph construction for use in image classification. This paper also details how use of this context data may reduce the need for manual annotation as well as increase the accuracy of crowd-sourced datasets. The results of their paper were a notable 27% improvement over some more traditional crowdsourced dataset such as the one used in Schmidhuber, Meier, and Ciresan's *Multi-column deep neural networks for image classification*[Wan+12].

Very Deep Convolutional Networks for Large-Scale Image Recognition

by Karen Simonyan and Andrew Zisserman

This paper is very similar to Schmidhuber, Meier, and Ciresan's *Multi-column deep neural networks for image classification* in regards to the use of small convolutions in deep layers for accurate neural network riamge recognition. Using the ImageNet 2014 challenege and two others, they demon-

strate both state-of-the-art accuracy on each dataset, as well as scalability and decent result to processing time efficiency. The creators of this paper have also made their bes convolutional network models publicly available, so our project may use them as a model for our own networks. Since their models are relatively efficient and simple, they are great candidates for testing out and comparing out gathered datasets with their manually annoted counterparts. By using other's models, we save the time needed to design and revise the actual network models so that we may focus on conducting experiments as to the efficiency and effectiveness of crowdsourced versus manually annotated data [SZ14].

Semi-supervised Relational Topic Model for Weakly Annotated Image Recognition in Social Media

by Z. Niu et al.

This paper focuses on improving image recognition with weakly annotated tags by encoding tags as relations among images. They then used a semi-supervised relational topic model in order to create another more explicit data model, which was then used to train a nondiscriminative classifier without further data processing. The success of this paper indicates that there is certainly a wealth of data available fo ruse through social media, except that it may need tobe extracted and the format slightly altered in order to be used for successful training. Our project will attempt to understand which aspects of data pre-processing will be constructive for our dataset generator [Niu+14].

ImageNet Classification with Deep Convolutional Neural Networks

by Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton.

This paper describes the ImageNet classifier, its structure, its operation, and the results of a trained version of the model. The authors achieve state-of-the-art predictions on the ImageNet LSVRC-2010 contest, and discuss their use of dropout layers in an effort to prevent overtraining of their deep convolutional network on their 1.2 million image dataset. Our project uses this model as a reference for the operation of our own neural networks and methods for improving the resulting predictions from our networks [KSH17]

Building High-Level Features Using Large Scale Unsupervised Learning

by Quoc Le et al.

This paper discusses how to build and use the features that may be extracted by the regression in a large-scale network model. They do however use unlabeled data, which is the worst case scenario for our project, in that all the labels are mismatched. We derive some ideas from their methodology of handling the lack of labels and attempt to apply them to our model. The authors results do show that it is possible to train their face detection model without labels, which is an amazing feat, and could work in parallel to our tentatively labeled datasets [Le+12].

TensorFlow: A System for Large-scale Machine Learning

by Martin Abadi et al.

This paper focusses on constructing a code-base for machine learning via dataflow graphs. The released Tensorflow library that accompanied the paper offers fast and efficient functions for creating, training, and testing a machine-learning model. Our project used this library as its basis for all deep learning techniques via the python Tensorflow library. It should be noted that Tensorflow utilizes GPUs to massively increase the model speed, which is especially importannt for my project for which training a network takes on the order of a day via tensorflow optimizations, whereas without it would take on the order of months [Aba+16].

Python Reddit API Wrapper

by Bryce Boe

This library allows easy access to the Reddit API in Python. This is extremely important for my project as the dataset link are gathered from reddit before being gathered from wherever the images themselves are hosted. The Python Reddit API Wrapper allows for mass collection of Reddit posts via the post search by timestamps. PRAW also limits the number of requests sent to Reddit servers in order to abide by Reddit's Terms of Service (TOS) [Boe14].

References

[Aba+16] Martin Abadi et al. "TensorFlow: A System for Large-scale Machine Learning". In: *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*. OSDI'16. Savannah, GA, USA: USENIX Association, 2016, pp. 265–283. ISBN: 978-1-931971-33-1. URL: http://dl.acm.org/citation.cfm?id=3026877.3026899.

[Boe14] Bryce Boe. "PRAW: The Python Reddit API Wrapper". 2014. URL: https://github.com/praw-dev/praw.

[KSH17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Commun. ACM* 60.6 (May 2017), pp. 84–90. ISSN: 0001-0782. DOI: 10.1145/3065386. URL: http://doi.acm.org/10.1145/3065386.

[Le+12] Quoc Le et al. "Building high-level features using large scale unsupervised learning". In: *International Conference in Machine Learning*. 2012.

[ML12] Julian McAuley and Jure Leskovec. "Image Labeling on a Network: Using Social-Network Metadata for Image Classification". In: *Computer Vision – ECCV 2012*. Ed. by Andrew Fitzgibbon et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 828–841. ISBN: 978-3-642-33765-9. URL: https://cs.stanford.edu/~jure/pubs/image-eccv12.pdf.

[Niu+14] Z. Niu et al. "Semi-supervised Relational Topic Model for Weakly Annotated Image Recognition in Social Media". In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 4233–4240. DOI: 10.1109/CVPR.2014.539.

[SMC12] J. Schmidhuber, U. Meier, and D. Ciresan. "Multi-column deep neural networks for image classification". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*. Vol. 00. June 2012, pp. 3642–3649. DOI: 10.1109/CVPR.2012.6248110. URL: doi.ieeecomputersociety.org/10.1109/CVPR.2012.6248110.

[SZ14] Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *CoRR* abs/1409.1556 (2014). arXiv: 1409.1556. URL: http://arxiv.org/abs/1409.1556.

[Sze+15] Christian Szegedy et al. "Rethinking the Inception Architecture for Computer Vision". In: *CoRR* abs/1512.00567 (2015). arXiv: 1512.00567. URL: http://arxiv.org/abs/1512.00567.

[Wan+12] Sheng-Yuan Wang et al. "Learning by expansion: Exploiting social media for image classification with few training examples". In: *Neurocomputing* 95 (2012). Learning from Social Media Network, pp. 117 –125. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2011.05.043.