Classification and mapping Twitter images

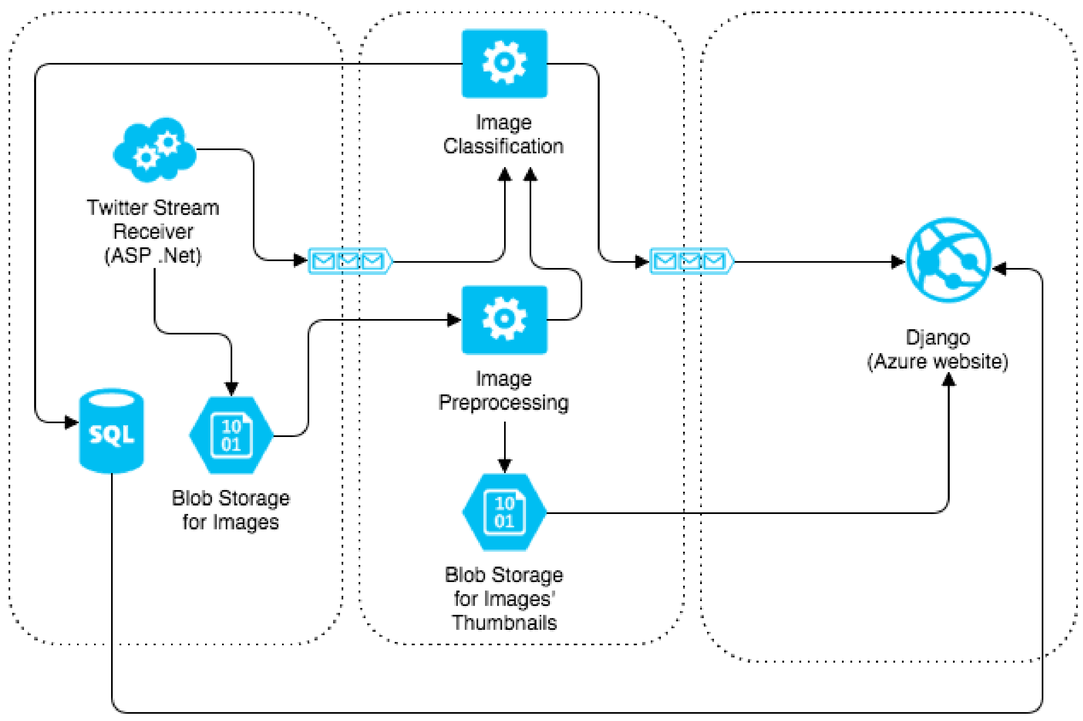
Andrey Poletaev[1], *Nikita Debelov*[2], *Maxim Ryabinskiy*[3]

[1]*Crystallnix, LLC, Omsk,*
[2]*Northern(Arctic)FederalUniversity, Arkhangelsk,*
[3]*Seismotech, LLC, Moscow*

Overview

Originally this project was built in a quite short period of time during Microsoft Research Russia Summer School "Doing Research in the Cloud" [4]. The main idea of this project is to collect, analyze and visualize data from social network Twitter. The first version was build on Microsoft Azure platform and heavily utilises it's tools such as: Queue Service, Blob storage service, Table service, SQL Azure Database, Bing Maps API. Parts of the system were launched on three trial accounts provided by Microsoft for school participants. The second version was build with using Celery, a distributed task queue – system to process messages, while providing operations with the tools required to maintain such a system. [1]. In both cases to use Twitter API – subscribe and receive tweets with both image and geotag, a Python library Tweepy [7] was used.

Architecture of first version



Web interface (first version)

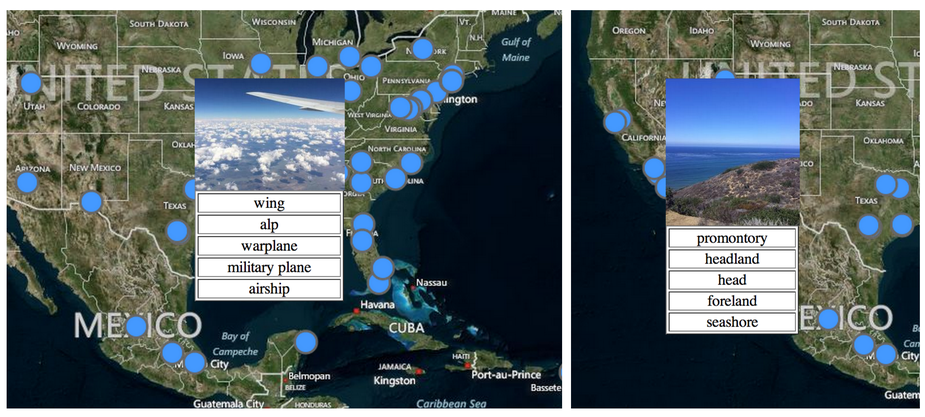
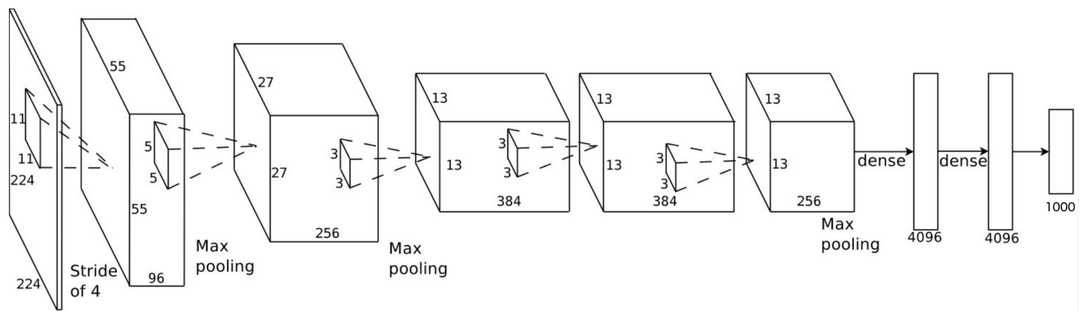
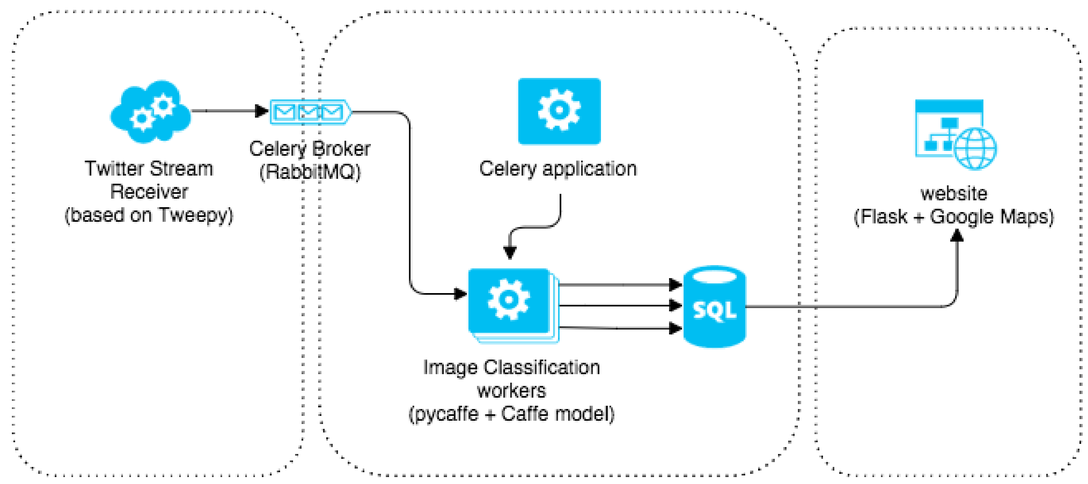


Image classification

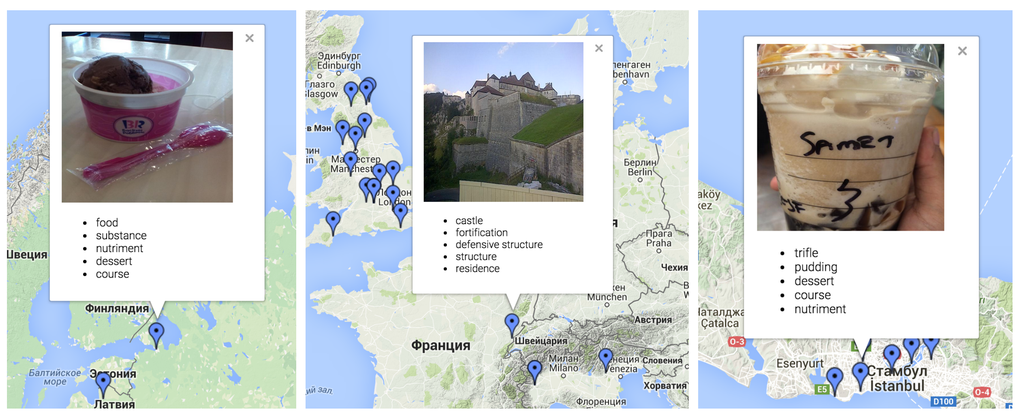
As classifier for images obtained from Twitter, deep convolutional neural network. Python library Caffe with prebuilt Caffe Reference ImageNet Model (implementation of an ImageNet model trained on ILSVRC-2012) were used.
Here's short overview of this model: the best validation performance during training was iteration 358,000 with validation accuracy 57.258% and loss 1.83948. This model obtains a top-1 accuracy 57.1% and a top-5 accuracy 80.2% on the validation set. [2] Max-pooling layers follow first, second, and fifth convolutional layers. The number of neurons in each layer is given by 253440, 186624, 64896, 64896, 43264, 4096, 4096, 1000. [3]



Architecture of second version



Web interface (second version)



References

[1] Celery – distributed task queue. Accessed: 2015-08-23.

[2] Yangqing Jia. Caffe: An open source convolutional architecture for fast feature embedding, 2013.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[4] Microsoft research russia summer school "doing research in the cloud". Accessed: 2014-08-18.

[5] E. Olivetti, S. M. Kia, and P. Avesani. MEG Decoding Across Subjects. *ArXiv e-prints*, April 2014.

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[7] Tweepy – python library for accessing the twitter api. Accessed: 2015-08-23.