Machine Learning Engineer Nanodegree

Capstone Proposal - Udacity

Isaac Ben-Akiva

April 13th, 2017

Intel & MobileODT Cervical Cancer Screening (Kaggle[1] Competition)

Domain Background and Motivation

Deep Learning is an emerging subfield of Machine Learning, which allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction (LeCun, Bengio, & Hinton, 2015). Deep Learning has made major advances in solving problems that have resisted the best attempts of the artificial intelligence community for many years. It has dramatically improved the state-of-the-art in areas such as speech recognition (Hinton & al, 2012), image recognition (Krizhevsky, Sutskever, & Hinton, 2012) and natural language processing (Collobert, et al., 2011). Deep Learning has beaten other machine-learning techniques at predicting the activity of potential drug molecules (Ma & et al, 2015), analysing particle accelerator data (Ciodaro, Deva, Seixas, & Damazio, 2012), reconstructing brain circuits (Helmstaeder et al., 2013), and predicting the effects of mutations in non-coding DNA on gene expression and disease (Xiong & al, 2015).

Computer Vision has become ubiquitous in our society, with applications in search, image understanding, apps, mapping, medicine, drones, and self-driving cars. Core to many of

---

[1] https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening

these applications are visual recognition tasks such as image classification, localization and detection. Recent developments in Deep Learning approaches have greatly advanced the performance of these state-of-the-art visual recognition systems. Convolutional Neural Network (aka CNN or ConvNet) is the state-of-the-art Deep Learning algorithm, which is often used in image recognition. It has achieved the highest accuracy in many image recognition challenges such as ImageNet, MNIST and NORB databases. ConvNets were inspired by biological processes and are variations of multilayer perceptrons designed to use minimal amounts of pre-processing (LeCun, Bottou, Bengio, & Haffner, 1998).

According to recent statistical data published by the World Health Organisation (WHO), cervical cancer is the fourth most common cancer in women, and the seventh overall, with an estimated 528,000 new cases in 2012. There were an estimated 266,000 deaths from cervical cancer worldwide in 2012, accounting for 7.5% of all female cancer deaths. However, cervical cancer, if caught is the pre-cancerous stage, is relatively easy to treat. Thus, the problem is to correctly identify a cervix in a pre-cancerous stage and the cervix type to administer the appropriate treatment. By training a Machine learning algorithm to classify the cervix from images, we may help healthcare providers by classifying the cervix type and what cancerous stage it is in. As we have shown above, deep learning has achieved a very high accuracy in image recognition. Therefore, it is natural to think we can apply this technique to help to prevent human death.

My personal motivation for investigating this area more in depth is that I believe Machine Learning is a powerful tool that can be used to improve human lives, especially in healthcare.

Problem Statement

Cervical cancer is so easy to prevent if caught in its pre-cancerous stage that every woman should have access to effective, life-saving treatment no matter where they live. Today, women worldwide in low-resource settings are benefiting from programs where cancer is identified and treated in a single visit. However, due in part to lacking expertise in the field, one of the greatest challenges of these cervical cancer screen and treat programs is determining the appropriate method of treatment which can vary depending on patients' physiological differences. Especially in rural parts of the world, many women at high risk for cervical cancer are receiving treatment that will not work for them due to the position of their cervix. This is a tragedy: health providers are able to identify high risk patients, but may not have the skills to reliably discern which treatment which will prevent cancer in these women. Even worse, applying the wrong treatment has a high cost. A treatment which works effectively for one woman may obscure future cancerous growth in another woman, greatly increasing health risks.

Currently, MobileODT offers a Quality Assurance workflow to support remote supervision which helps healthcare providers make better treatment decisions in rural settings. However, their workflow would be greatly improved given the ability to make real-time determinations about patients' treatment eligibility based on cervix type.

In this project, we will use Deep Residual Network (He & et al, 2015) to classify accurately a woman's cervix type based on cervical images. Correctly identifying the cervix type is essential to prevent ineffectual treatments and allow healthcare providers to give proper referral for cases that require more advanced treatment.

Datasets and Inputs

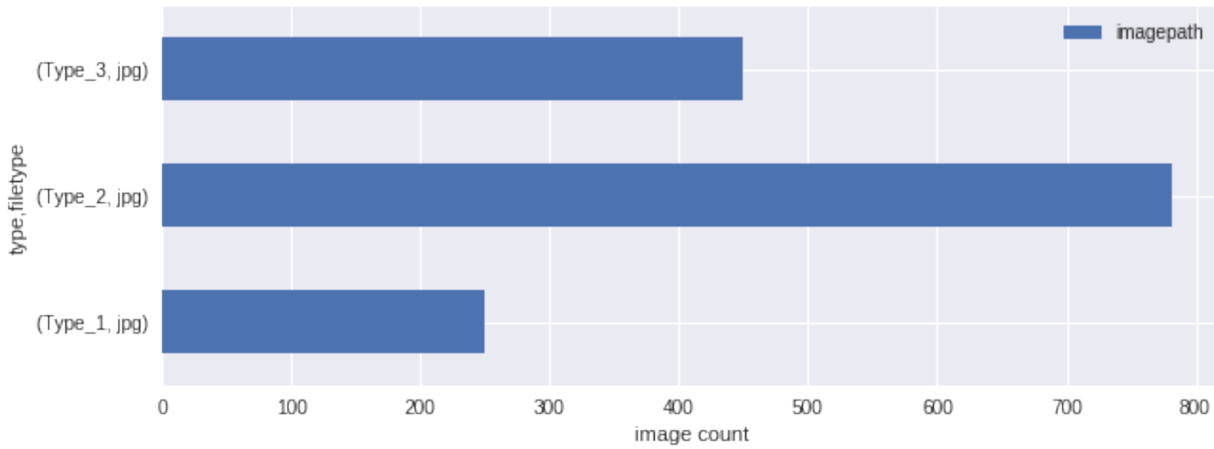
This project will use the following datasets provided by Intel and MobileODT on Kaggle's (https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/data) competition website:

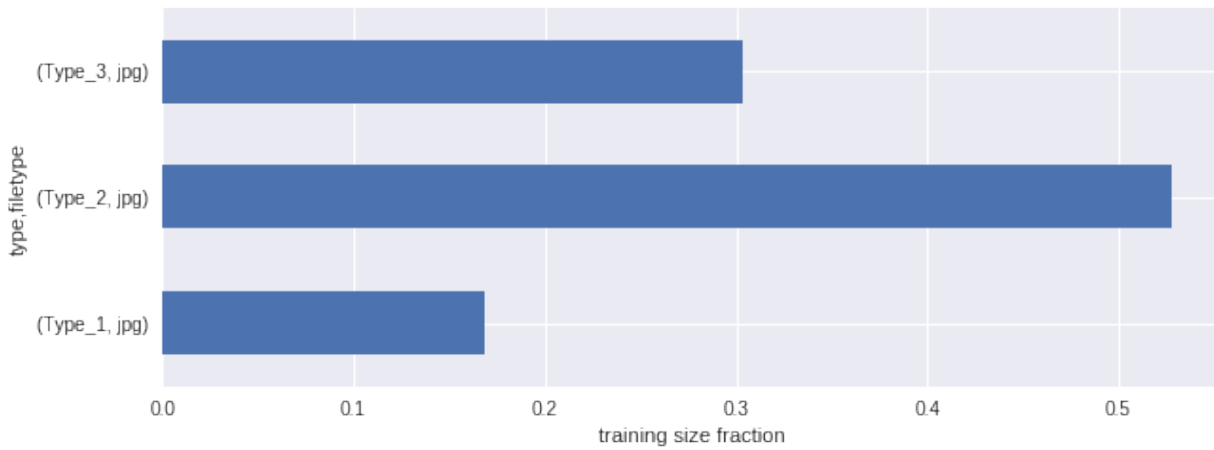
- **train.7z**[2] – the training set. The images are organised in their labelled categories: Type_1, Type_2 and Type_3.
- **test.7z**[3] – the first stage test set. All the images are in this same folder.
- **additional_Type_{x}.7z**[4] – Additional images to help training the models. These images sometimes come from duplicated patients, so the images might look alike since they are taken in the same session.

An initial exploratory data analysis reveals that all the files are in JPG format and Type 2 is the most common one with a little bit more than 50% of the training set, Type 1 is about 20% and Type 3 takes the remaining 30% of the training set as shown on the below figure.

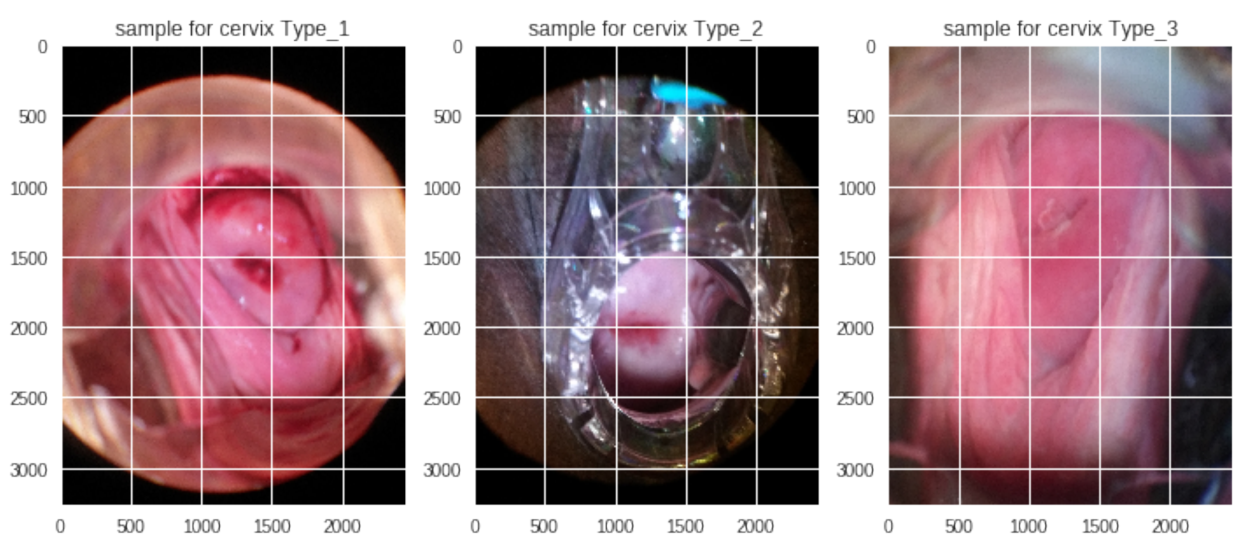
---

[2] https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/download/train.7z
[3] https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/download/test.7z
[4] https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/download/additional_Type_{x}_v2.7z





Now, let's have a look at each type to get an idea about how the images look like.



The images seem to vary a lot in their formats. The first two samples have only a circular area with the actual cervix in the centre. The third sample has the image in a rectangle.

Finally, let's have a look at the image dimensions to get an idea of how many different shapes of images by class there are. For that purpose, we will take a small sample of approximately 36 images per type.

	nchans	**ncols**	**nrows**	**type**	**0**
3	3	2448	3264	Type_3	28
6	3	3096	4128	Type_3	7
7	3	3264	2448	Type_3	1
5	3	3096	4128	Type_2	18
2	3	2448	3264	Type_2	17
8	3	4128	3096	Type_2	1
1	3	2448	3264	Type_1	18
4	3	3096	4128	Type_1	17
0	3	2322	4128	Type_1	1

As we can see, there are different image dimensions per class. We will need to rescale them down prior to feed them into our Convolutional Network.

Solution Statement

We propose the use of Convolutional Neural Networks. Our solution will be based on a pre-training ResNet model (He & et al, 2015). This is a promising solution as Deep Residual Network is the state-of-the-art in image classification in the ImageNet challenge.

Benchmark Model

For this problem, we will use the competition Leaderboard score on Kaggle as a benchmark for the model. We will benchmark against the current leading score of 0.52216 established by user DataGeek. We will try to reduce the gap between algorithm and his until the end of the competition on Kaggle.

Evaluation Metrics

The solution will be evaluated using the multi-class logarithmic loss function. Each image has been labelled with one type. For each image, we will submit a set of predicted probabilities, one for every category. The formula is then,

$$logloss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log\left(p_{ij}\right),$$

where N is the number of images in the test set, M is the number of categories, log is the natural logarithm, y_{ij} is 1 if observation i belongs to class j and 0 otherwise, and p_{ij} is the predicted probability that observation i belongs to class j. It's worth noting that according to the competition webpage on Kaggle, the submitted probabilities for a given image are not required to sum to one because they are rescaled prior to being scored, each row is divided by the row sum. In order to avoid the extremes of the log function, predicted probabilities are replaced with $\max(\min(p, 1-10^{-15}))$.

Project Design

We will take the following approach to solve the problem posed by this project:

1. Exploratory Data Analysis
2. Pre-processing and Images preparation
3. Model Development
4. Model train/train

5. Model evaluation

The Deep Learning solution will be a Residual Network (ResNet) with 50 layers developed with Keras and TensorFlow backend. The model will be trained on the following hardware:

- Ubuntu 16.04
- x64 i7 Processor with 32GB RAM
- GPU GeForce GTX 1070 6GB
- 2TB HDD

Programming Language and Libraries

- Python 2.7
- scikit-learn – open source machine learning library for Python
- Tensorflow 1.0 – open source library for deep learning
- Keras – open source neural network library written in Python. It’s capable of running on top of either TensorFlow or Theano
- OpenCV – open source library for computer vision

References

Ciodaro, T., Deva, D., Seixas, J. M., & Damazio, D. (2012). Online particle detection with Neural Networks based on topological calorimetry information. *Journal of Physics: Conference Series, 368*, 12-30.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research, 12*, 2493-2537.

Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jain, V., Seung, H. S., & Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 168-174.

Hinton, G., & al, e. (2012, October 18). Deep neurla networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine, 29*(6), pp. 82-97.

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. P. Weinberger (Ed.), *Advances in Neural Information Processing Systems 25* (pp. 1097-1105). Curran Associates, Inc.

LeCun, Y., Bengio, Y., & Hinton, G. (2015, May 28). Deep learning. *Nature, 521*, 436-444.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278-2324.

Mikolov, T., Deoras, A., Povey, D., Burget, L., & Cernocky, J. (2011). Strategies for training large scale neural network language models. *IEEE Workshop on Automatic Speech Recognition and Understanding* (pp. 196-201). IEEE.

Mitchell, T. (1997). *Machine Learning.* McGraw-Hill Science.

Xiong, H. Y., & al, e. (2015, January 9). The human splicing code reveals new insights into the genetic determinants of disease. *Science, 347*(6218).