
Classify Subcellular Protein Patterns In Human Cells

Yixiao Song, Dongkai Chen
Department of Computer Science
Stanford University

Abstract

This paper evaluates three types of convolutional neural network architecture: simple convolutional neural network, Xception, GapNet on the human proteins dataset. The uniqueness of datasets are 4 channels of protein pictures. GapNet has better performance on classifying the human proteins images with better F1 score.

1 Introduction

Proteins execute their function at specific times and locations within a cell. Understanding the location of proteins in cell organelles is fundamental to understand biological processes [5]. Researchers previously have been limited to the single protein patterns. In order to understand the intricacy of human cells, we develop deep neural networks to classify the mixed patterns of proteins. The neural network will be able to identify protein types and protein's location from the high throughput images.

Convolutional Neural Networks (CNNs) are state-of-the-art techniques for image classification in a variety of fields, with variants like ResNet[3], InceptionV3 [8], and Fully Convolutional Network. The CNNs increase performance significantly and have become the dominant approach to object detection. The drawback of convolutional neural networks is that internal data representation of a convolutional neural network does not take into account critical spatial hierarchies between simple and complex objects.[7]

On top of InceptionV3 [8], We also implement Xception [6] which have proven a significant outperforms InceptionV3. And GapNet-PL [1] which able to process high resolution image.

2 Dataset and Features

The dataset is from Kaggle competition. A scaled set of 512x512 PNG files in train.zip and test.zip is available. There are total 28 different labels in the data set. All image samples are represented by four channels(green, blue, red, yellow). According to the competition rule, the green filter should be used to predict the label, and the other filters are used as references. The four-channel images may be different from imagenet, so it is challenging to handle the transfer learning. So we divide the yellow channel data by 3 then combine 1/3 of yellow channel to the RGB images in the milestone stage.

One of the challenges from the dataset is the significant massive data imbalance, such as "Nucleoplasm" are in majority portion of the dataset, while there are many rare classes, like "Endosomes," "Lysosomes," and "Rods & rings." It is crucial to augment the rare dataset set or choose a suitable loss function.

Due to the size of the original picture (a mix of 2048x2048 and 3072x3072 TIFF files around 250GB), we are using a scale down version(512x512 PNG 18GB) for initial implementation. With data argumentation, we decide to reduce the size further down to speed up learning & validation cycle.

Based on 50 set of random examples, human can easily distinguish the difference between the samples with 299*299 version.

Latter for GapNet-PL [1], we are able to use 700*700 version as input which shows relatively large improvement to the F1 score of validation set. Due to the resource limitation ,we didn't test 1024*1024 as the original paper processed. But we tests the 1024*1024 on simply convolutional neural networks.

3 Example of Data

From the dataset, the distribution of each category is unbalanced. "Nucleoplasm" and "Cytosol" have significant larger volume compare to the rest 25 classes. Below are the pictures which have different resolution

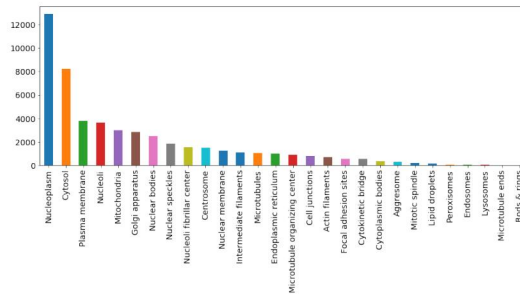


Figure 1: Distribution of the dataset

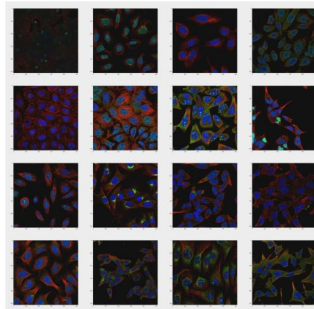


Figure 2: Example of Proteins with 512 by 512 scale

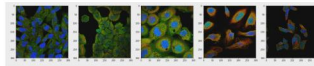


Figure 3: Example of Proteins with 299 by 299 scale

As shown below the train dataset is 80% and 20% is for the validation set.

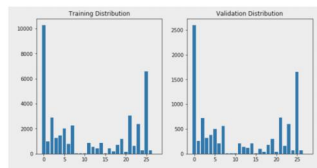


Figure 4: Data Distribution Between Train and Validation

4 Literature review

Based on the requirement, it fits into the category of image classification. The base model is a convolutional neural network with max pooling. One of the state-of-the-art methods for this task is called Xception [6].

4.1 Xception Networks

A convolutional neural network architecture based entirely on depthwise separable convolution layers. In effect, with the mapping of cross-channels correlations and spatial correlations in the feature maps of convolutional neural networks can be entirely decoupled. Because this hypothesis is a stronger version of the hypothesis underlying the Inception architecture.[6]

4.2 GapNet-PL

A new proposed architecture called GapNet-PL [1] can deal with high-resolution images and therefore has the possibility to learn from fine structures within images as they do not have to be downscaled.

GapNet-PL using two-step approach. Firstly , it it bundle few convolutional layer to represent a encoder and interspersed with max-pooling to learn abstract feature.

In a first step we use an encoder consisting of several convolutional layers, some with a stride of two, interspersed with max-pooling layers to learn abstract features on different spatial resolutions.

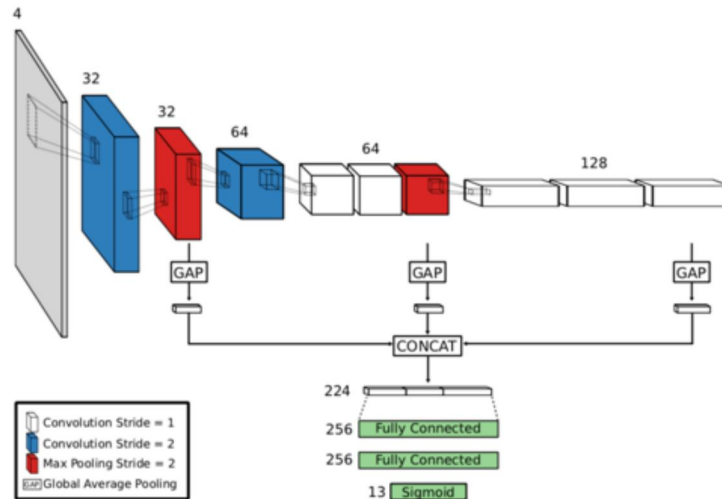


Figure 5: GapNet-PL as proposed in original paper

5 Approach

5.1 Data Augmentation

The uniqueness of this project are the color channels of input image , which are red, yellow,green and blue. Unlike any of the existed pre-trained model (eg, Inception Net) which input image have only 3 color channels (RGB). Therefore, we proposed 3 data argumentation methods and corresponding architectures.

5.1.1 Four channels input without argumentation process

Four input channels fully connected neural network can leverage all the information provided by the task. However, as stated before, there isn't any existed pre-trained model(parameters) we can use for transfer-learning.

5.1.2 Three channels with all the information

From chromatology perspective, yellow color consists of RGB. For example pure shades of yellow = $(r, g, b)(255, 255, 0)$. Therefore channel yellow can merge to channel red and green

As above data argumentation process, all the information will still be preserved and able to leverage existed pre-trained model for transfer learning.

5.1.3 Three channels without all channel yellow

Based on the above point, all the possible color can be generated with the combination of RGB. Adding color yellow to channel red and green could also have a negative impact on the model by the emphasis the weight on the color channels red and green. Therefore, the third proposal for data augmentation process is to remove the channel yellow from the input.

5.2 Simple Convolutional Neural Network

We implement a simply convolutional neural networks with two convolutional layers as a base model. The accuracy for 299x299x3 was 0.20 on training data after first epoch but after we switch to 1024x1024x4 images, the results are pretty well the accuracy was 0.944 on training data and 0.940 on validation data. The neural networks are only trained with 10 epoch.

5.3 Xception + Transfer Learning

The Xception architecture [6] is more accurate than previous state of the art models, as shown in the chart below, The Xception architecture shows a much larger performance improvement on the JFT dataset. Also we

We utilize the pre-trained two InceptionResNetV3 model from Keras and use three channels (RGB) + 3 channels (black and white) as shown in Figure 7 to train two separate model and multiply the two output to fully connected layer. (figure 8)

which gives an improvement on F1 score. (Figure 10)

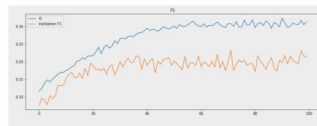


Figure 6: Xception F1 score

5.4 GapNet-PL

With the GapNet-PL architecture, we were able to train higher definition picture. Due to the hardware limitation we are not able to process 1024*1024 therefore we resize to 700*700, but still the F1 score improved compare to previous implementation(figure 11). [2]

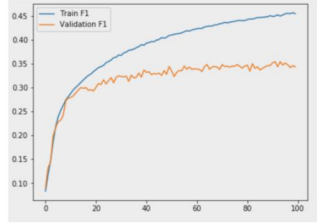


Figure 7: GapNet-PL F1 score

6 Evaluation

The F1 score shows the transfer learning from Xception has significant improvement in classifying protein than the simply convolutional neural network does. GapNet has better performance than the Xception and SNN. The image size is essential for the accuracy and loss by comparing the results in the table below. The 1024x1024 surprisingly perform very well on simply ConvNets with weighted loss function. A higher resolution image is the key to the neural networks learning the image features.

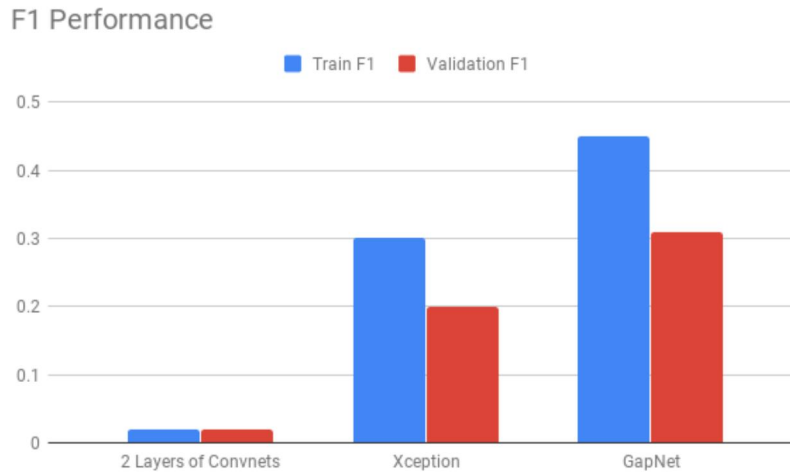


Figure 8: F1 score

	Train Acc	Validation Acc	Train Loss	Validation loss
Simply Convnets(RGB)(299x299x3)	0.122	0.121	0.8	0.78
Simply Convnets(RGBY)(1024x1024x4)	0.942	0.9	0.22	0.21
Xception(RGB)(299x299x3)	0.99	0.96	0.1	0.02
GapNet(RGBY)(700x700x4)	0.942	0.94	0.0219	0.0218

7 Conclusion/Future Work

This is a very challenge project. We would like to continue improving the model by implement SELU activation function[4] for the GapNet-PL, instead of Rectified Linear Unit (ReLU) with Batch normalization. It will reduce the memory cost and may allow to input higher definition pictures on GapNet. The F1 loss function is a valuable research topic, which may improve the neural networks on the performance of the F1 score.

8 Github Code Location

<https://github.com/chedy028/CS230-Final-Project>

9 Contributions

Yixiao Song and Dongkai Chen worked together on building the neural networks and writing this report. Both of us have equal contributions.

10 Code Reference

<https://www.kaggle.com/byrachonok/pretrained-inceptionresnetv2-base-classifier>

<https://www.kaggle.com/kmader/transfer-learning-for-human-protein-submission>

<https://github.com/kwotsin/TensorFlow-Xception>

<https://www.kaggle.com/kwentar/two-branches-xception-lb-0-3>

<https://www.kaggle.com/rejpalcz/gapnet-pl-lb-0-385>

<https://github.com/cedrickchee/capsule-net-pytorch>

References

- [1] P. u. d.-b. r. Anonymous authors. Human-level protein localization with convolutional neural networks. 2018, 2018.
- [2] M. Haltuf. Gapnet-pl [lb 0.385]. 2018.
- [3] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [4] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter. Self-normalizing neural networks. In *Advances in Neural Information Processing Systems*, pages 971–980, 2017.
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [6] X. Liu, F. Xu, Y. Sun, H. Zhang, and Z. Chen. Convolutional recurrent neural networks for observation-centered plant identification. *Journal of Electrical and Computer Engineering*, 2018, 2018.
- [7] S. Sabour, N. Frosst, and G. E. Hinton. Dynamic routing between capsules. In *Advances in Neural Information Processing Systems*, pages 3856–3866, 2017.
- [8] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.