

# ML2018FALL Final Proposal

## 1. 隊名及隊員

隊名：

NTU\_r06922134\_台北大冒險

隊員：

葉沛陽 r06922134

湯梵平 r06922120

廖彥綸 b05902001

林文韶 b05901157

## 2. 所選擇的題目

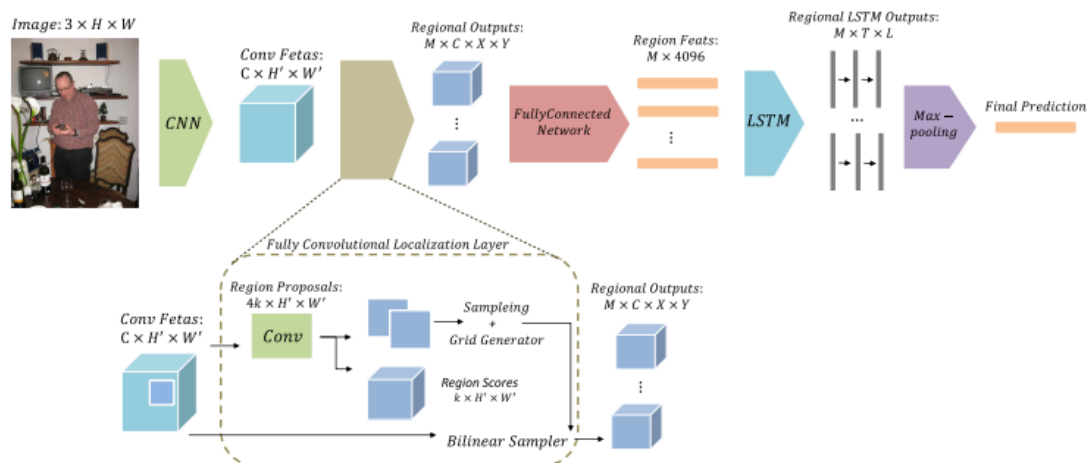
Human Protein Atlas Image Classification

## 3. problem study:

題目: Multilabel Image Classification With Regional Latent Semantic Dependencies

### 1. Regional Latent Semantic Dependencies (RLSD) model

使用 Regional Latent Semantic Dependencies (RLSD) model，考慮到除了 label visual 會影響分類 semantic dependencies 也會影響，所以這篇 paper 另外使用了 RNN 來抓 labels 之間的關聯。



### 2. Convolution Feature 部分:

使用 VGGNet 包含 13 個 convolution layers(  $3 \times 3$  kernel sizes) 和 五個  $(2 \times 2)$  max-pooling layer,然後使用 ReLU 跟 下圖的 loss function。

$$L(b, g) = \sum_{i \in x, y, w, h} \text{Smooth}_{L_1}(b_i, v_i)$$

$$\text{Smooth}_{L_1} = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise.} \end{cases}$$

### 3. Box sampling and bilinear interpolation :

使用 subsampling 來降低資料大小再傳進 LSTM. 在 test 階段，使用 non-maximum suppression 來挑選 the top M highest ranked proposals.

### 4. Fully-Connected Network:

用兩層 4096 維 fully-connected layers 跟 dropout.

### 5. Max-Pooling:

使用 Max-Pooling 因為較 average-Pooling 適合用來消除預測噪音， the fusion layer 的 output 被餵進 a multi-way softmax layer with the squared loss as the cost function, which is defined as:

$$J = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^L (p_i^j - g_i^j)^2$$

where y is ground truth and p is prediction.

### 6. Pre-train 方面:

- localization layer is pre-trained on the Visual Genome region caption dataset
- the LSTM is first pre-trained on the global image without region proposals, where every time step has the global image label as ground truth to compute loss as the initial of RLSD.

### 7. 參數方面:

label embedding size is 64.

one-layer LSTM(memory cell size is 512).

Optimization 用 SGD 以及 learning rate 為 0.00001.

### 8. 結論:

RLSD 非常適合預測小物品以及場景中物品有緊密關聯的影像。

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8310600&fbclid=IwAR3TOZVV5QphImkUPBcIZSgp3EU224RVebJvdfqOtLNLHkYywtYf2OACsgU&tag=1>

## 4. proposed method:

目前初步將 RGBY 四個檔案當作四個 channels，直接讀圖檔 512\*512 沒有 resize。

下圖為目前使用的 model 架構，主要使用一層 CNN 加上 keras 上 pretrained 好的 Inception\_resnet\_v2 架構，最後再加上一層 28 output 的 Dense，分類是採用 softmax，threshold 取 0.05。

Layer (type)	Output Shape	Param #
input_3 (InputLayer)	(None, 512, 512, 4)	0
batch_normalization_3 (Batch Normalization)	(None, 512, 512, 4)	16
conv2d_2 (Conv2D)	(None, 256, 256, 3)	15
inception_resnet_v2 (Model)	(None, 1536)	54336736
dense_1 (Dense)	(None, 28)	43036
Total params: 54,379,803		
Trainable params: 54,319,251		
Non-trainable params: 60,552		
None		

目前對於 RGBY 四個檔案直接使用，之後可能嘗試對他們作前處理，transformation 和 normalization。

目前採用 softmax 之後也打算改為 sigmoid 試看看有無改善。

目前也有發現資料有很嚴重的 Imbalanced 問題，之後可能要想辦法處理，目前想到的是針對每個 label 給予不同 class weight。

也打算針對每個 label 給予不同的 threshold，看看結果有無改善。

