09.28 MLEGLAB group meeting memo
by Jing Jin

1. Paper presenter： Yuxin Cui -noise to nosie paper: Learning Image Restoration without Clean Data

   This paper provide an algorithm learning through the noisy images to get a clean image. The key is that they only use the noisy data without the clean data in the whole training process. But there is one restriction. Even they add different noise to clean data to get different training sample, the original image should be the same one.

   The important statistic fact in many image restoration tasks(eg. The image with gaussian noise), is the expectation of the corrupted input data is the clean target, which derive from minimizing the L2 loss. As a result, when we are mapping dataset A with different noise to the dataset A with some other noise, we can get the minimal MSE when we actually map the dataset A with noise to clean dataset A. And this is the basic idea that they can get the clean data from training the noisy data.

   In some other image restoration tasks(e.g text noise), the noise is incremental but not continuous. The median of the corrupted input data is the clean data, which derive from minimize the L1 loss.

2. Research presenter: Zhonghao Liu- a DCNN model for HLA-peptide binding affinity prediction

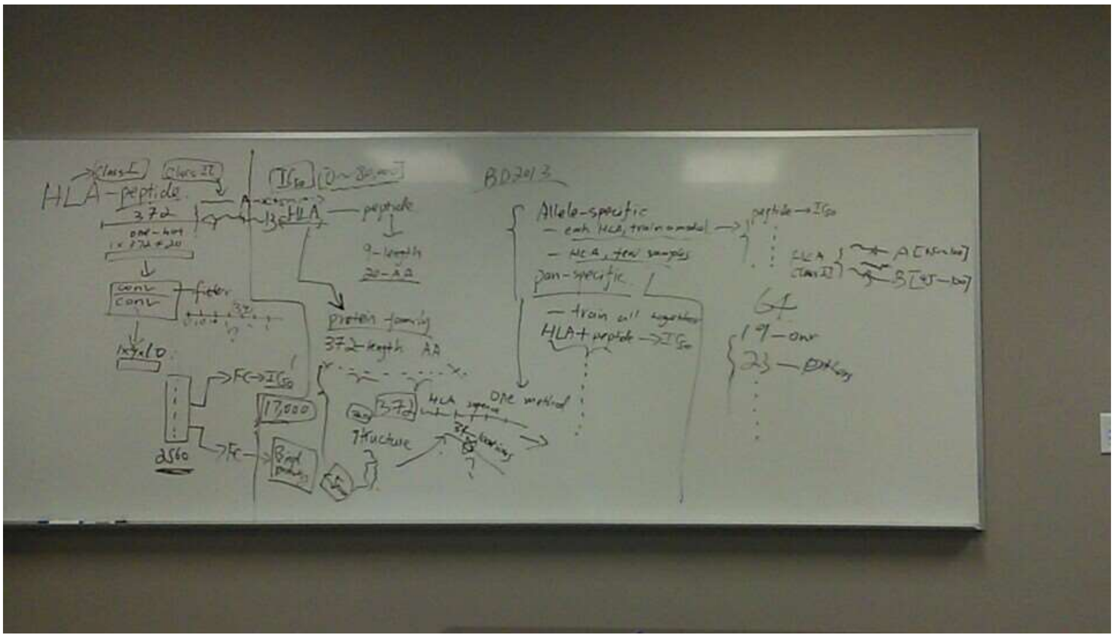
   This paper proposes a DCNN model to predict the class I HLA-peptide binding affinity. The model only needs raw sequences from the HLA-peptide binding pairs without structure information. The DCNN model consists of three parts.

   1) Peptide and HLA encoders: The peptide raw input tensor is 1*9*20, and the HLA raw input tensor is 1*372*21. There are two encoders separately converting the peptide and the HLA into two tensors with a unified dimension 1*9*10. Then the two tensors are concatenated to get a encoded feature tensor with dimension 1*9*20. This will be fed in to the next part-the binding context extractor.
   2) Binding context extractor: The extractor takes into the encoded feature tensor from 1) and outputs a 2560-dimension vector, which is actually the binding context between a peptide and a HLA. Especially, this extractor uses locally connected layers to retain the position related information.
   3) Affinity and binding predictors: Another novel design of this model is that both the binding probabilities and the IC50 Value(used to define the binding affinity) are used as the output in the final stage. The author argues that the classification predictor can force the network to predict a more accurate IC50 values.

   Questions:

   Dr.Hu: use the algorithm to evaluate the critical 34 locations in the HLA mentioned in another article, or we can get new meaningful positions of HLA sequence based on this algorithm.

   Pictures:



HLA-peptide
372
Conv
Conv
filter
2560
FC
IC50
17,000
HLA
peptide
9-length
20-AA
protein family
372-length AA
structure
BD2013
Allele-specific
— each HLA, train a model
— HLA, few samples
pan-specific
— train all together
HLA+peptide → IC50
one method