



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Preprocessing, Data Augmentation and Feature Extraction

Gabriel Simões



BUSINESS INTELLIGENCE AND
MACHINE LEARNING RESEARCH GROUP

Agenda

- Preprocessing
- Data Augmentation
- Feature Extraction

Preprocessing

Boring stage where we should observe:

- Duplicities
- Correct labeling
- Corrupted data
- Outliers
- Train, test and validation splits
- Normalization



Preprocessing

Let's get some practice in our sandbox!

Data Augmentation



375
500



375
500



128
312

Data Augmentation



375
500



375
500



128
312

Data Augmentation



375
500



375
500



128
312



78% smaller!

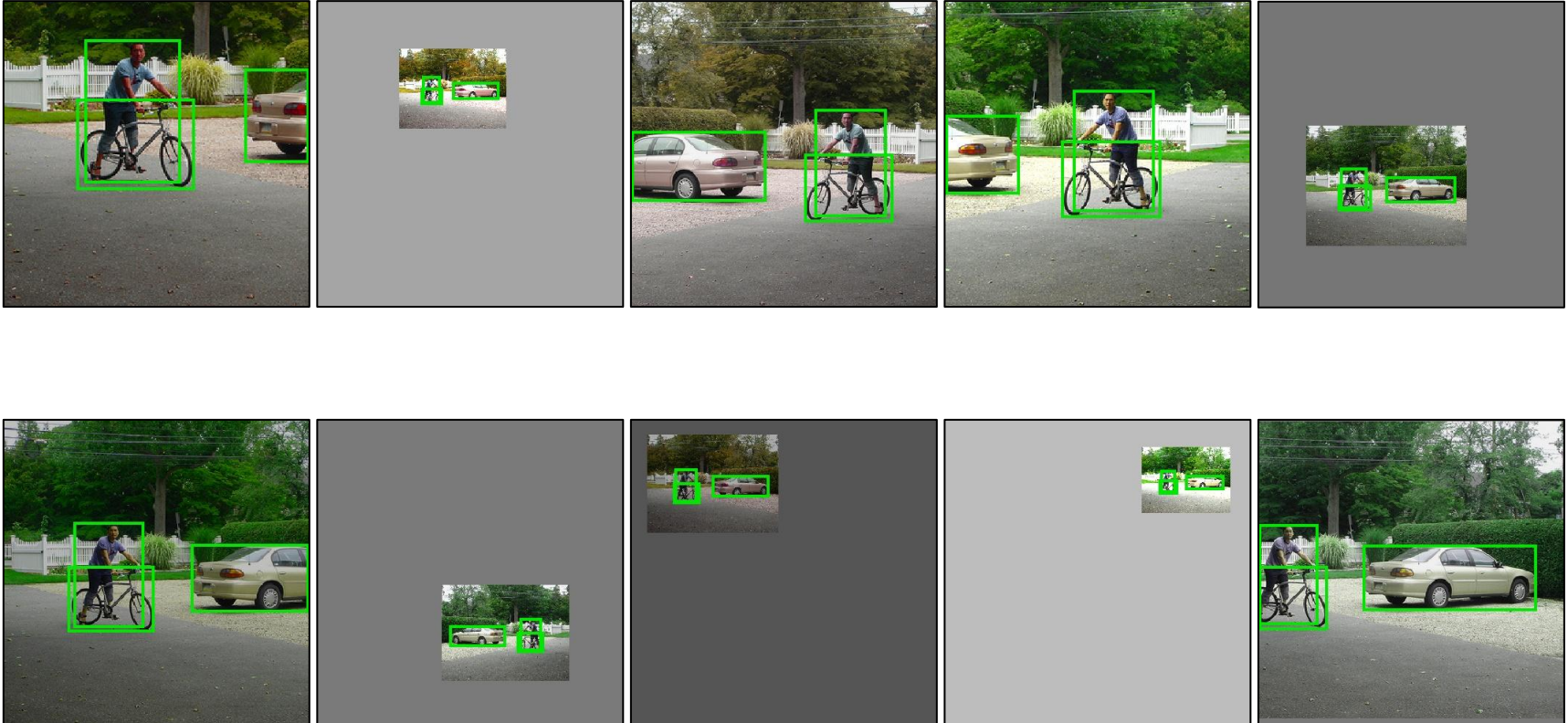
Data Augmentation



416

416

Data Augmentation



416

416

Data Augmentation

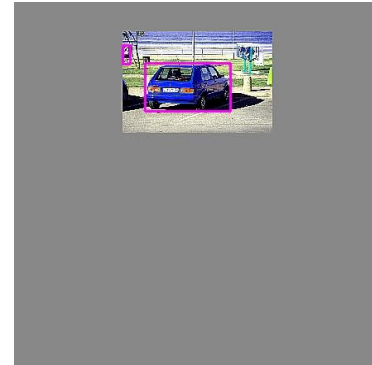
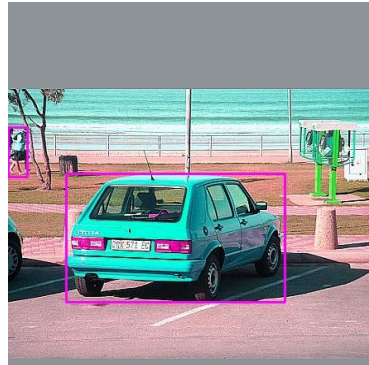


416

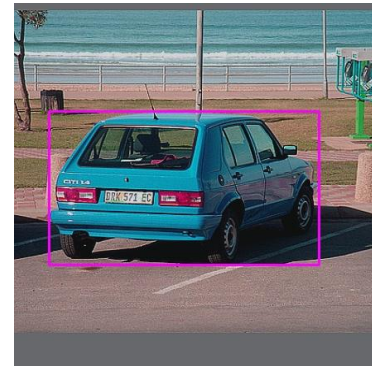
416

Data Augmentation

Scale: 0.25 to 2.00



Jitter: 0.2



brightness, contrast
and saturation

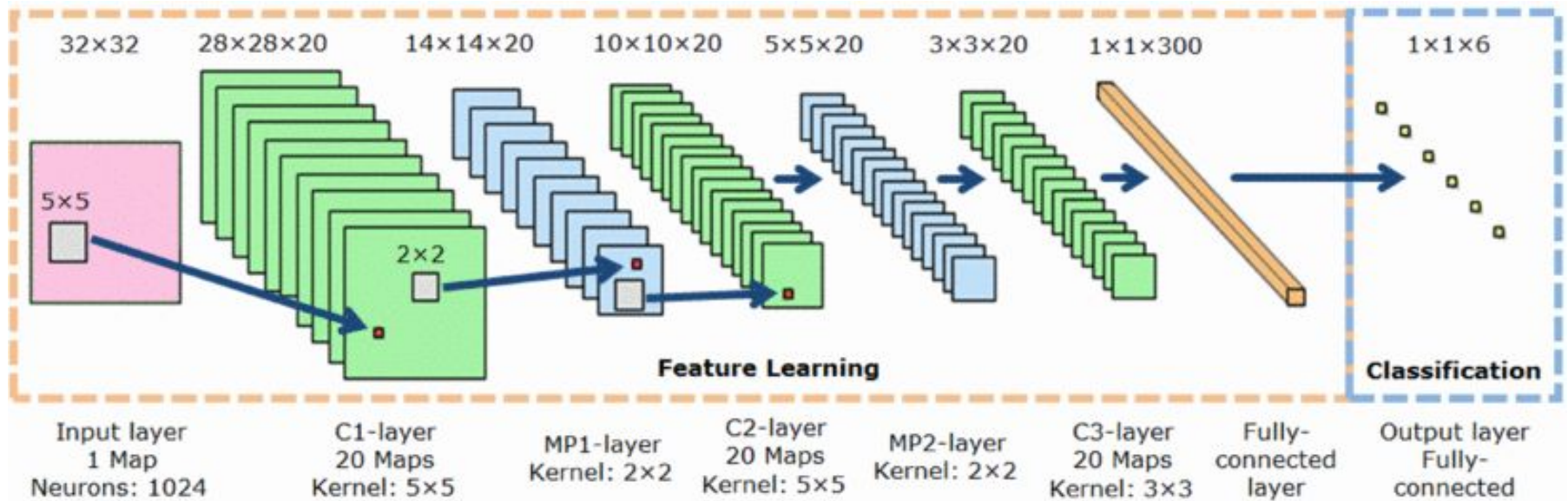


Data Augmentation

Let's get some practice in our sandbox!

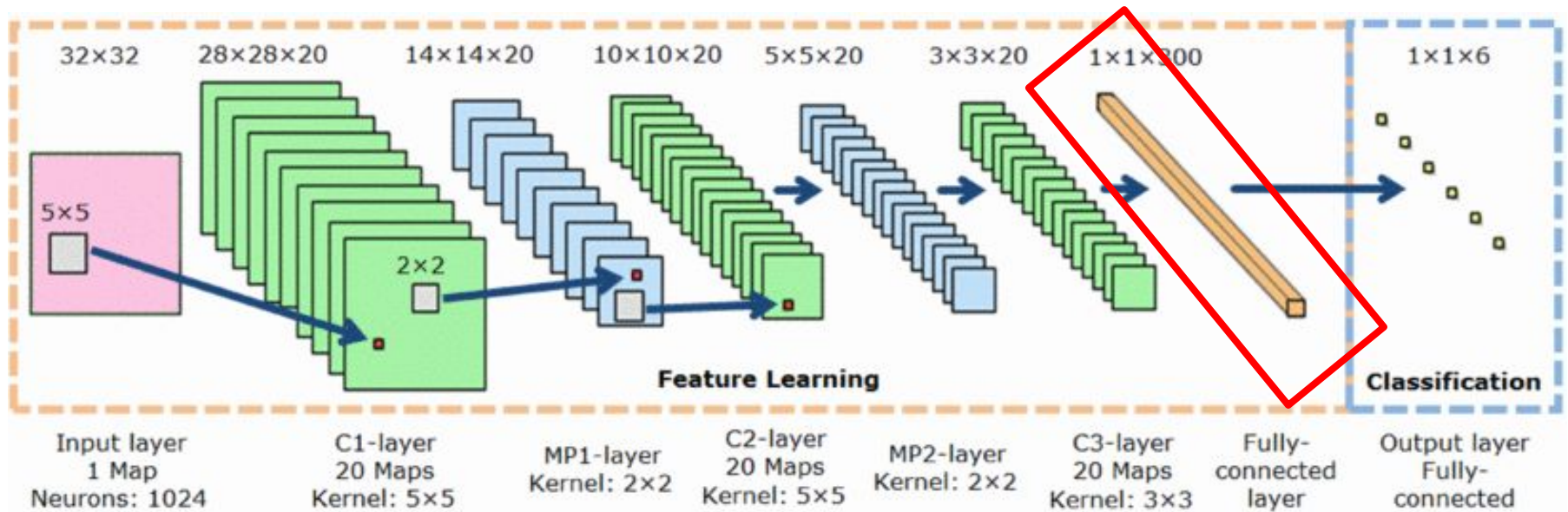
```
git clone https://github.com/Pezaun/image_augmentation.git
```


Feature Extraction



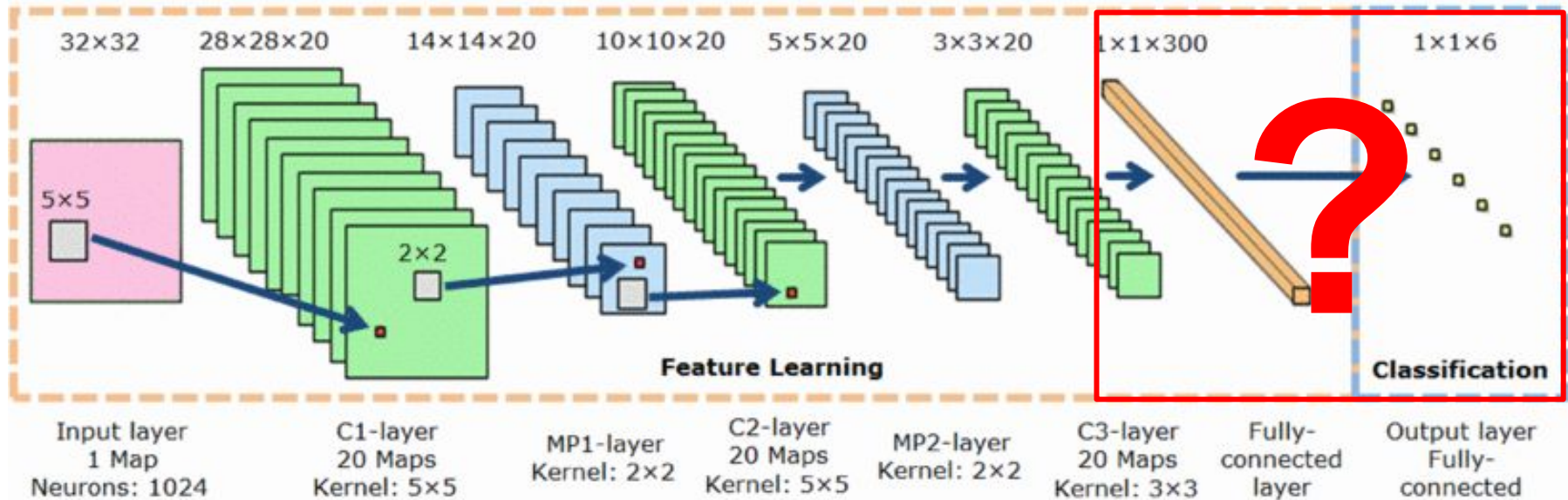
Assume a classifier based on a convolutional network trained to classify cat, dog, horse, cow, duck and frog. You have an excellent 32×32 image classifier for 6 classes!

Feature Extraction



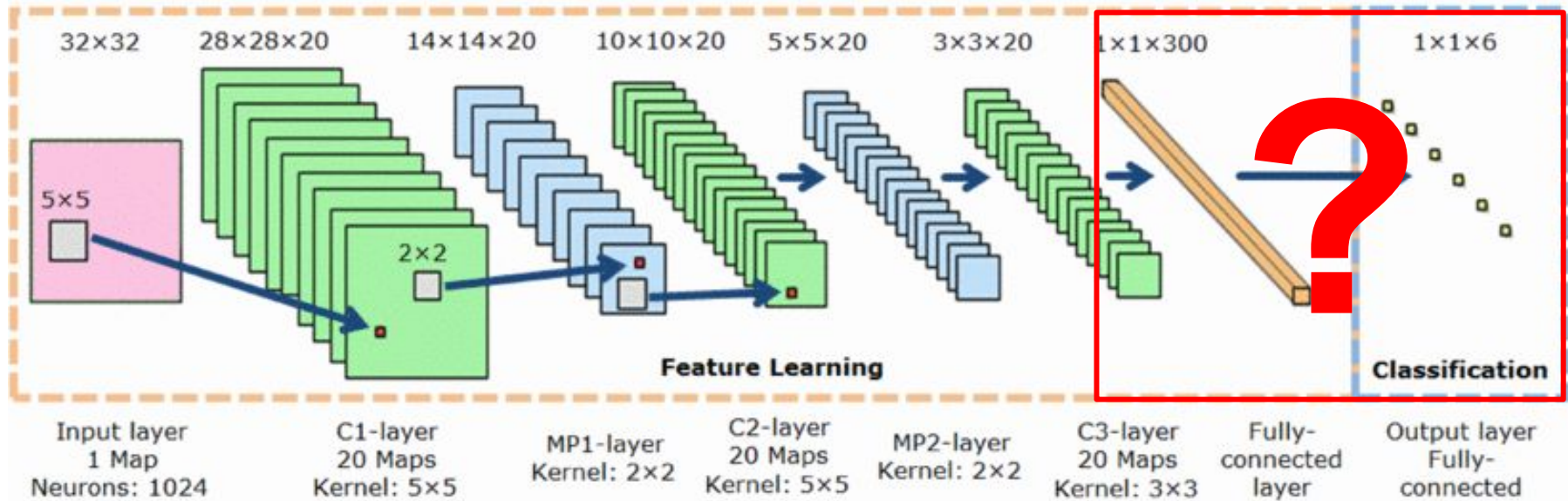
Recall that the classifier is defined by the last layer, usually composed of a number of neurons compatible with the number of classes.

Feature Extraction



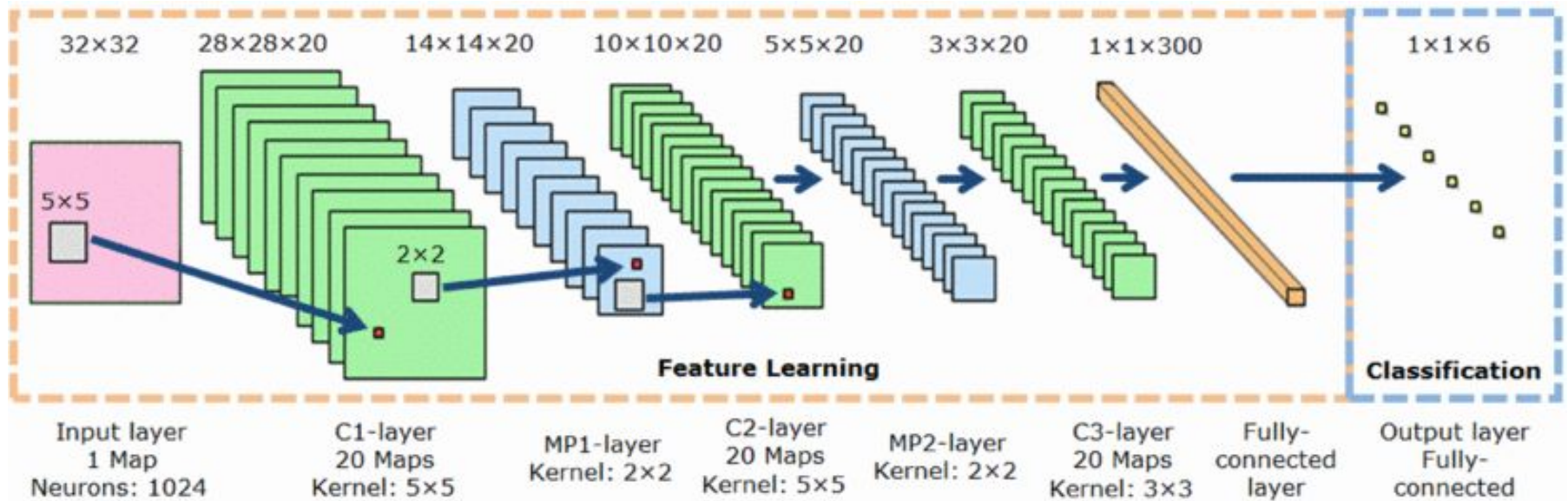
Now assume that you need to modify this classifier to classify only cat, dog and horse. Your new problem has only 3 classes! What do you do?

Feature Extraction



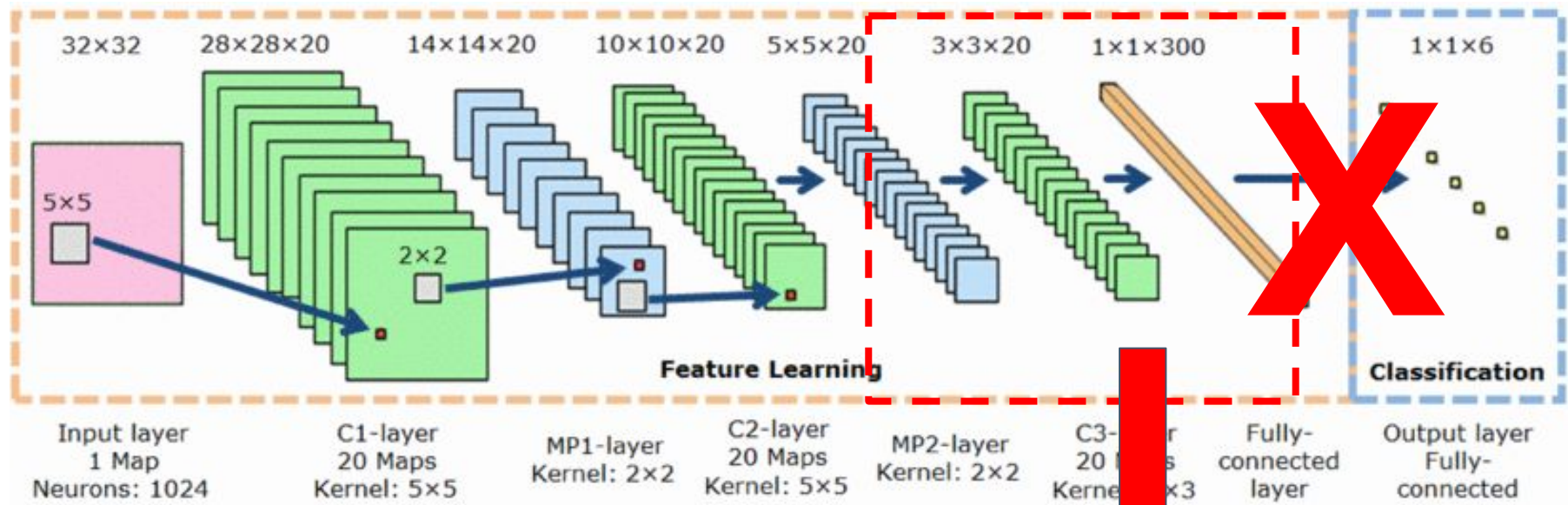
We have 3 simple options! Today, we will focus on feature extraction.

Feature Extraction

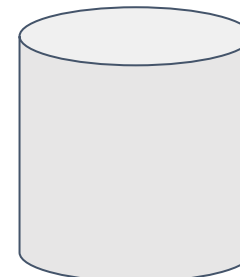


If your network knows the concepts you need, it must know how to generate features that describe these concepts well.

Feature Extraction



So, forget the classifier and just keep the features!



Feature Extraction

Let's get some practice in our sandbox!

git clone https://github.com/Pezaun/feature_extraction.git