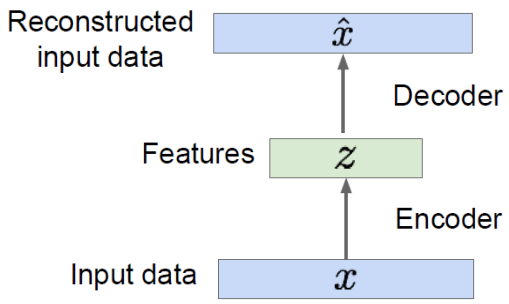
CENG 506 Deep Learning
Exercise Sheet

1. **10 points.** After tuning your model, you find that softmax classifier works well. Specifically, the last layer of your network computes scores for each class and they are fed into softmax function. The model achieves 100% accuracy on the training data. However, you observe that training loss does not reach zero. Tell why softmax (cross-entropy) loss can never be zero?

2. **10 points.** What's the risk with tuning hyperparameters using a test dataset?

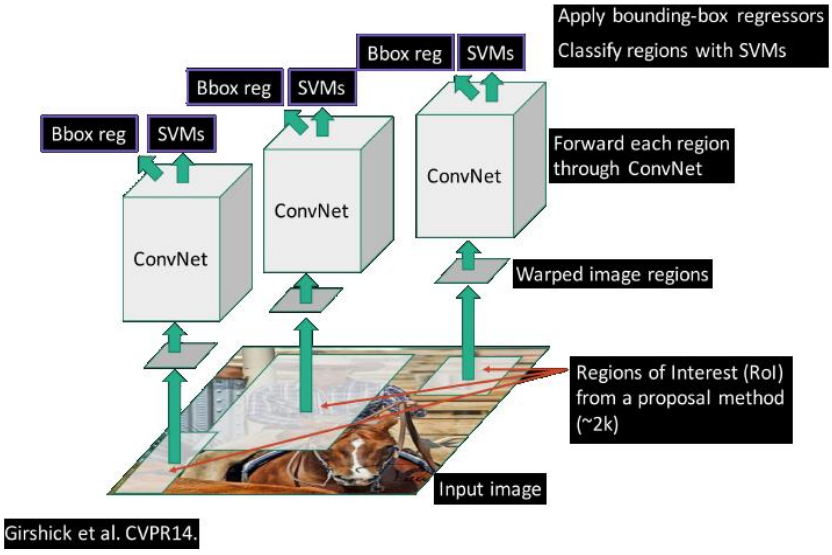
3. **15 points.** We want to give a city landscape image to a NN and get back the same image but this time it should look like it is taken at night time. What would be the strategy to train such a network? Hint: Autoencoders and GANs

4. **15 points.** In autoencoders, we have a decoder network that converts feature vector (z) to a reconstructed input ($\hat{x}$). If our input is an image and if we use CNN for the decoder network, explain how z is converted to $\hat{x}$? How a vector becomes an image by CNN?



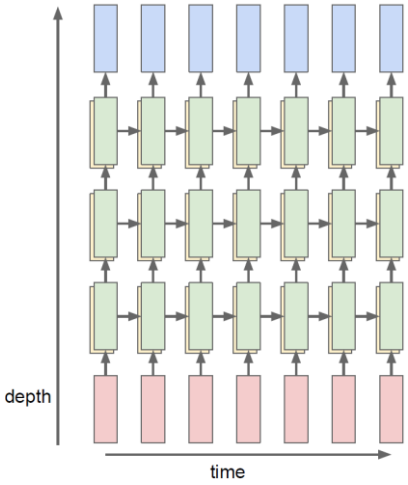
5. Why we do not update NN weights after seeing each sample and wait for seeing a batch of samples like 32, 64 or 100?

6. R-CNN pipeline for object detection is given in the image below.



   a. Explain the two main reasons why this method is quite slow.
   b. Explain how Fast R-CNN solves these two issues.

7. An ‘unrolled’ multi-layer RNN is given below. Please explain (give mathematics) how do we obtain a hidden layer unit from previous and lower layer hidden layer units?



8. How we can benefit from transfer learning if we have a large dataset of new classification examples. You want to perform image classification with a separate dataset consisting of some animals taken by different people (touristic photos). But this dataset is relatively small, not enough to effectively train a deep neural network. How would you do training?
Hint: A CNN pre-trained on ImageNet dataset is available for your use.

9. When the input is 2-dimensional, you can plot the decision boundary of your neural network and clearly see if there is overfitting. How do you check overfitting if the input is 10-dimensional?

10. In RNNs we are obliged to take input sequentially one by one, which is time consuming. But with transformers we can take input as a whole. Still, for a model it is important to know which word comes before/after another one. How transformers know about the location of words in a sentence? Explain the procedure.