1. Write any two differences between RNN and CNN?

CNN	RNN
a neural network that is specialized for processing a grid of values **X** such as an image	a neural network that is specialized for processing a sequence of values $x(1), \ldots, x(\tau)$
Process images with large width and height	Process longer sequences
Mostly process images of fixed size	Most recurrent networks can process sequences of variable length.
output will not dependent on previous members of the output	Each member of the output is a function of the previous members of the output

2. What are the challenges of long term dependencies.?
   Vanishing and exploding gradients
3. Give a real world application of visual attention
   Google Street view
4. What are the learning models used for creating adversarial Networks.

    Generative models, discriminative models

5. What is Echo state network?

Echo state network is a type of Recurrent Neural Network, part of the reservoir computing framework, which has the following particularities:
the weights between the input -the hidden layer (the 'reservoir') : Win and also the weights of the 'reservoir': Wr are randomly assigned and not trainable ,the weights of the output neurons (the 'readout' layer) are trainable and can be learned so that the network can reproduce specific temporal patterns .The hidden layer (or the 'reservoir') is very sparsely connected (typically < 10% connectivity)

6. what is vanishing and exploding gradient descent?
   - Recurrent networks, construct very deep computational graphs by repeatedly applying the same operation at each time step of a long temporal sequence.
   - Repeated application of the same parameters gives rise to especially pronounced difficulties.
   - During training, we compute the prediction error $E^{(k)}$ and use the Back Propagation Through time algorithm to compute the gradient

$$\frac{\partial E}{\partial W} = \sum_{t=1}^{T} \frac{\partial E_t}{\partial W}$$

The gradient of the error term in an RNN

If we have learning task that includes large time steps, the derivative expression **tends to vanish when time stpes are large,** due to the derivative of the tanh activation function which is smaller than 1.
The product of derivatives can also explode if the weights are large enough to overpower the smaller tanh derivative, this is known as the exploding gradient problem.