

Introduction to Deep Learning

Assignment Project Exam Help

Angelica Sun

(adapted from Atharva Paudelkar, Jingbo Yang)

Add WeChat powcoder



Overview

- Motivation for deep learning
- Convolutional neural networks
- Recurrent neural networks
- Transformers
- Deep learning tools

Assignment Project Exam Help

<https://powcoder.com>

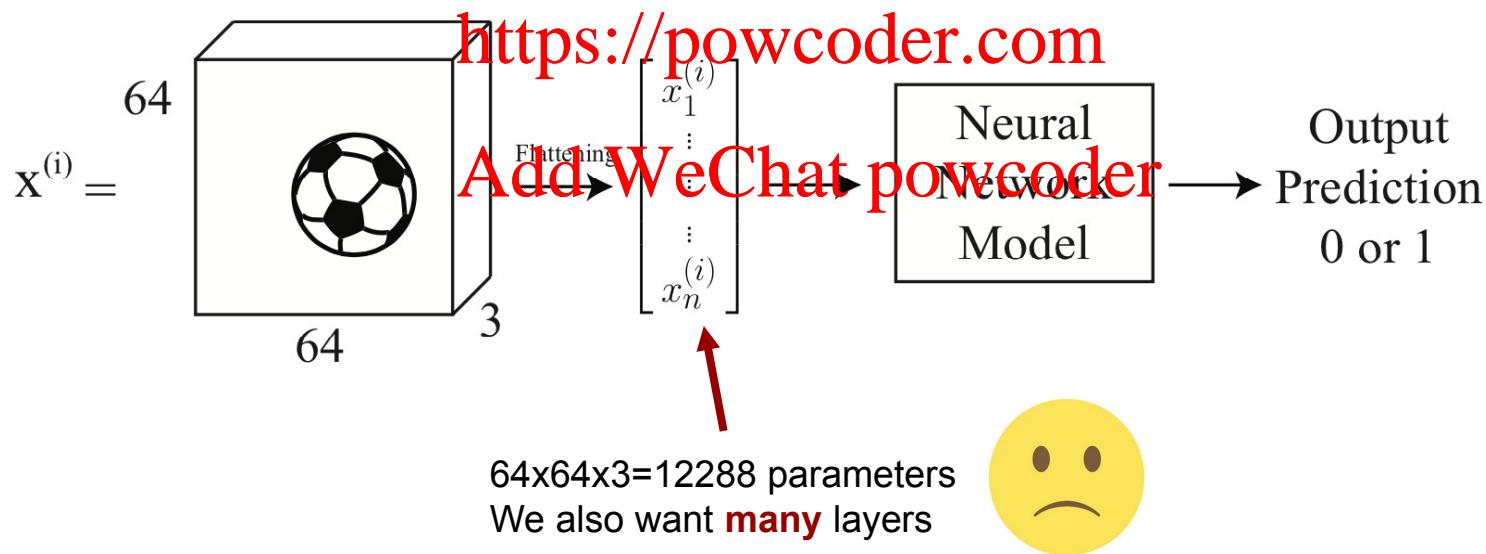
Add WeChat powcoder

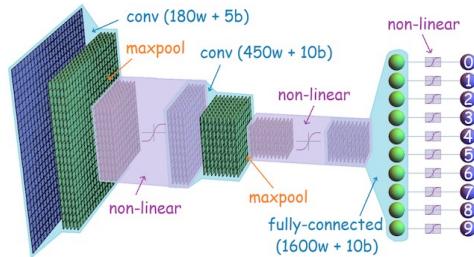
But we learned multi-layer perceptron in class?

Expensive to learn. Will not generalize well.

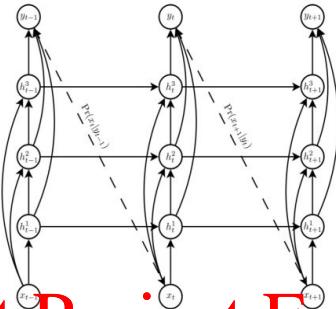
Assignment Project Exam Help

Does not exploit the order and local relations in the data!



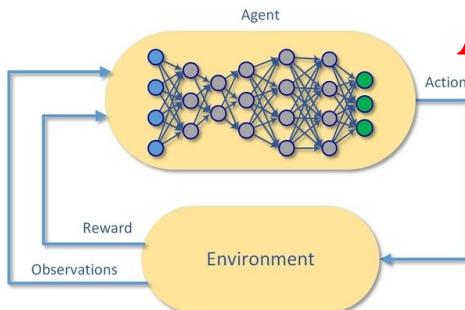


Convolutional NN
Image

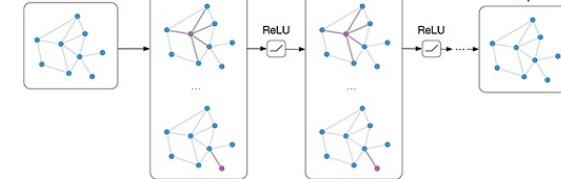


Recurrent NN
Sequential Inputs

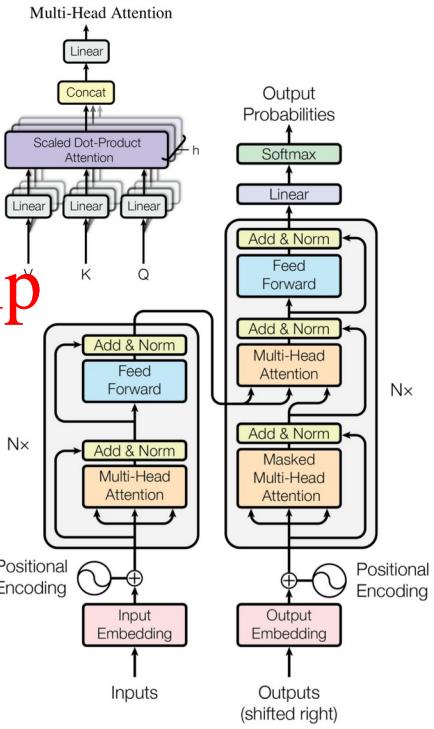
What are areas of deep learning?
<https://powcoder.com>



Deep RL
Control System



Graph NN
Networks/Relational



Transformers
Parallelized
Sequential Inputs

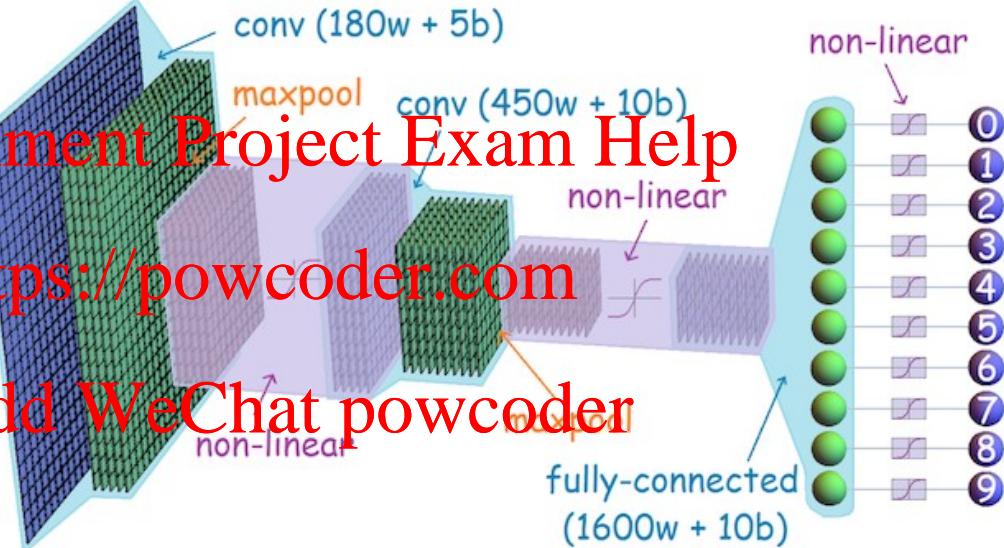
Starting from CNN

Convolutional
Neural Network

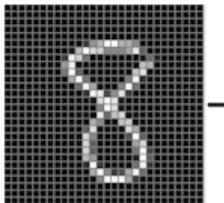
Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Let us look at images in detail

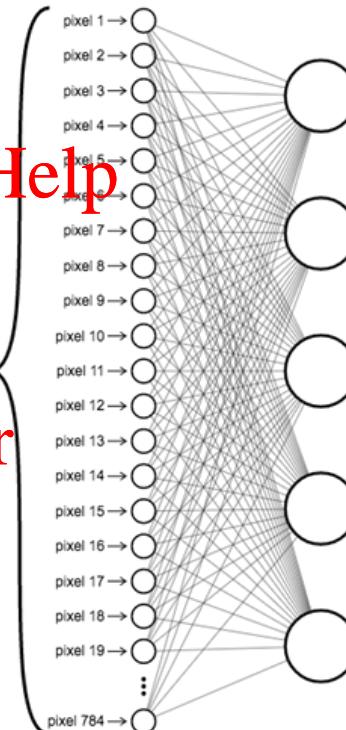


28 x 28
784 pixels

Assignment Exam I
<https://powcoder.com>
Add WeChat powcode

Assignment Project Exam Help

<https://powcoder.com>

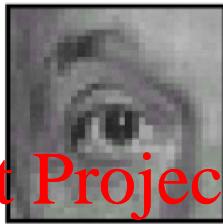


Filters in traditional Computer Vision



Original

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$



Assignment Project Exam Help

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

https://powcoder.com
Add WeChat powcoder



$$*(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} - \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}) =$$



Sharpening filter
(accentuates edges)



Input

Image credit:

<https://home.ttic.edu/~rurtasun/courses/CV/lecture02.pdf>

Learning filters in CNN

Why not extract features using filters?

Better, why not let the data dictate
what filters to use?

Learnable filters!!

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



1 _{x1}	1 _{x0}	1 _{x1}	0	0
0 _{x0}	1 _{x1}	1 _{x0}	1	0
0 _{x1}	0 _{x0}	1 _{x1}	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

Convolved Feature

Convolution on multiple channels

Images are generally RGB !!

How would a filter work on a image
with RGB channels?

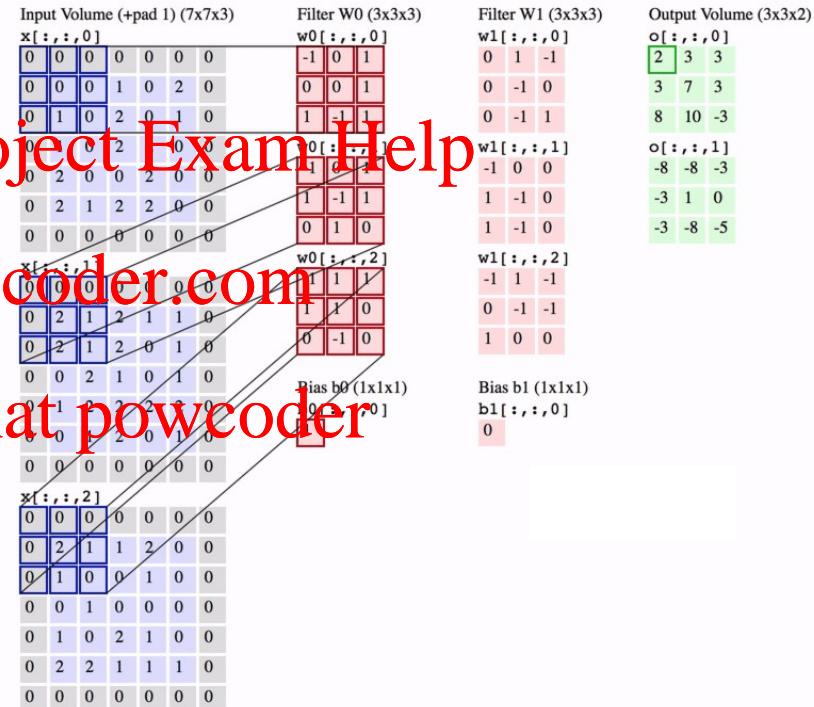
The filter should also have 3
channels.

Now the output has a channel for
every filter we have used.

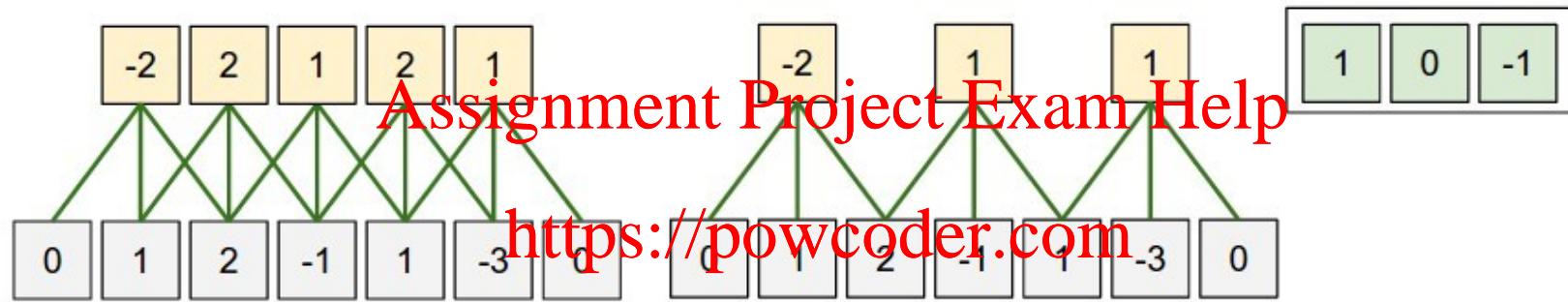
Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Parameter Sharing

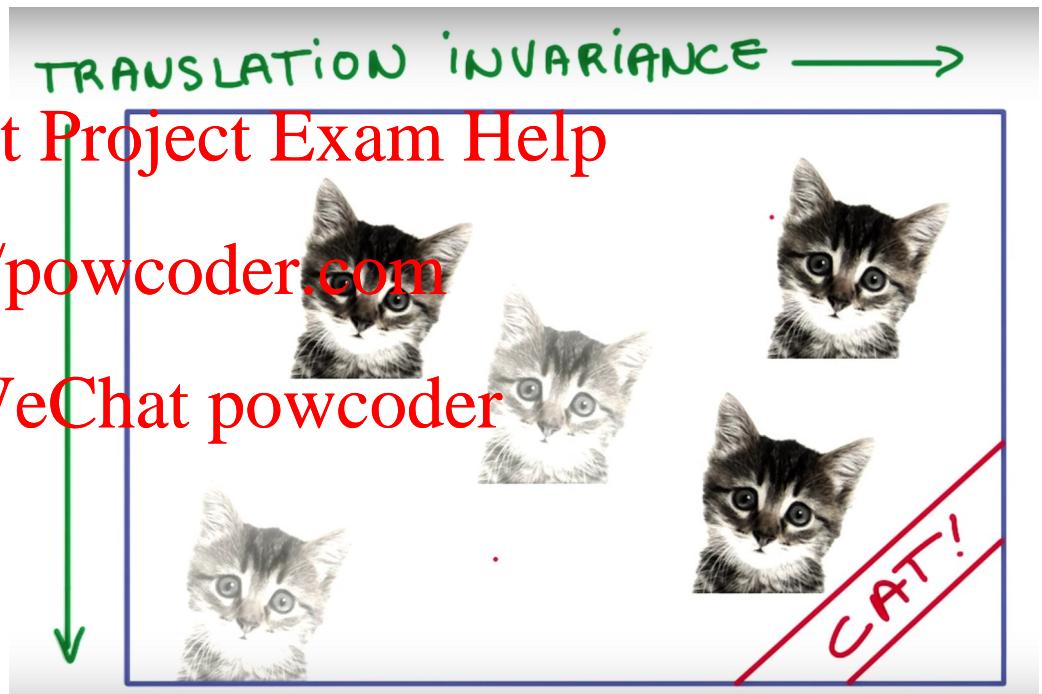


Lesser the parameters less computationally intensive the training. This is a win win as we are reusing parameters.

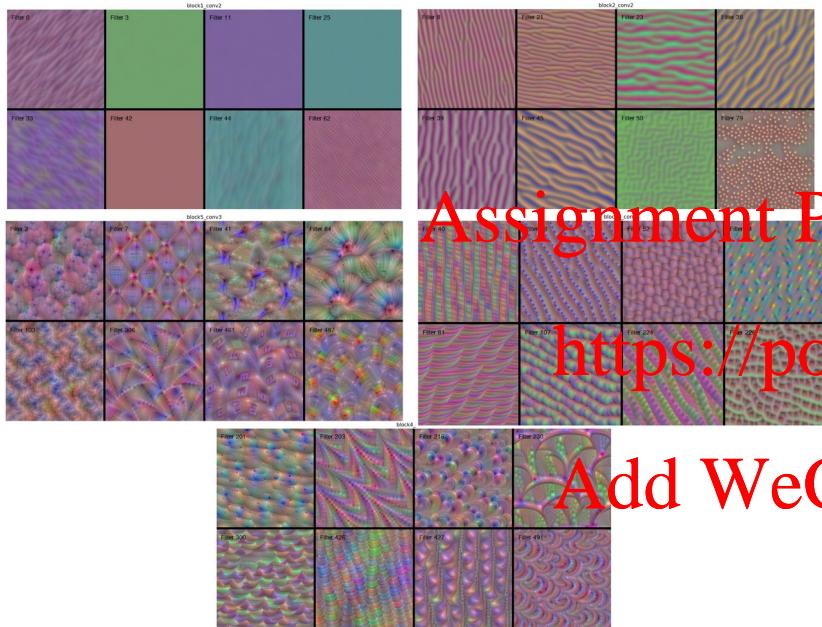
Translational invariance

Since we are training filters to detect cats and the moving these filters over the data, a differently positioned cat will also get detected by the same set of filters.

Add WeChat powcoder



Visualizing learned filters



Images that maximize filter outputs at certain layers. We observe that the images get more complex as filters are situated deeper

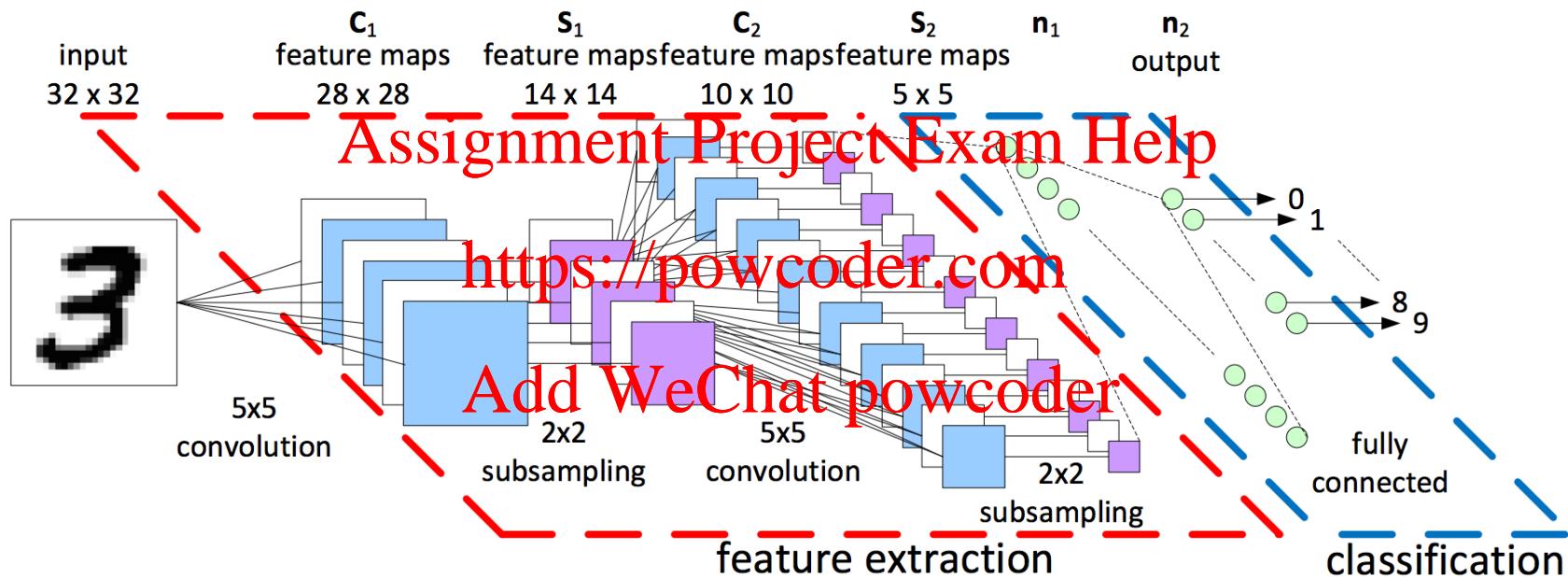
How deeper layers can learn deeper embeddings. How an eye is made up of multiple curves and a face is made up of two eyes.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

A typical CNN structure:



Convolution really is just a linear operation

In fact convolution is a giant matrix multiplication.

Assignment Project Exam Help

We can expand the 2 dimensional image into a vector and the ~~conv~~

<https://powcoder.com>

Add WeChat powcoder

$$\begin{pmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 \end{pmatrix} * \begin{pmatrix} k_1 & k_2 \\ k_3 & k_4 \end{pmatrix} = \begin{pmatrix} k_1 x_1 + k_2 x_2 + k_3 x_4 + k_4 x_5 \\ k_1 x_2 + k_2 x_3 + k_3 x_5 + k_4 x_6 \\ k_1 x_4 + k_2 x_5 + k_3 x_7 + k_4 x_8 \\ k_1 x_5 + k_2 x_6 + k_3 x_8 + k_4 x_9 \end{pmatrix}$$

SOTA Example – Detectron2



How do we learn?

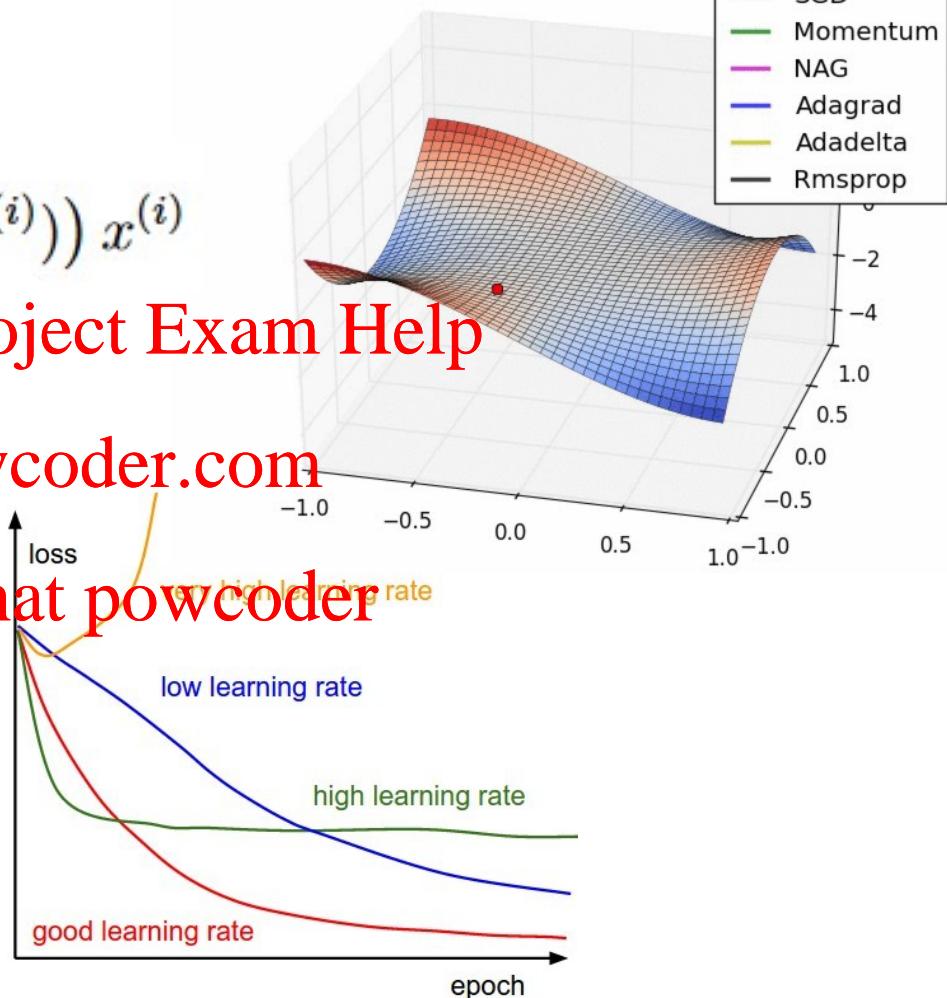
Instead of $\theta := \theta + \alpha (y^{(i)} - h_\theta(x^{(i)})) x^{(i)}$

They are “optimizers”

- Momentum: Gradient + Momentum
- Nestrov: Momentum + Gradients
- Adagrad: Normalize with sum of squares
- RMSprop: Normalize with moving avg of sum of squares
- ADAM: RMSprop + momentum

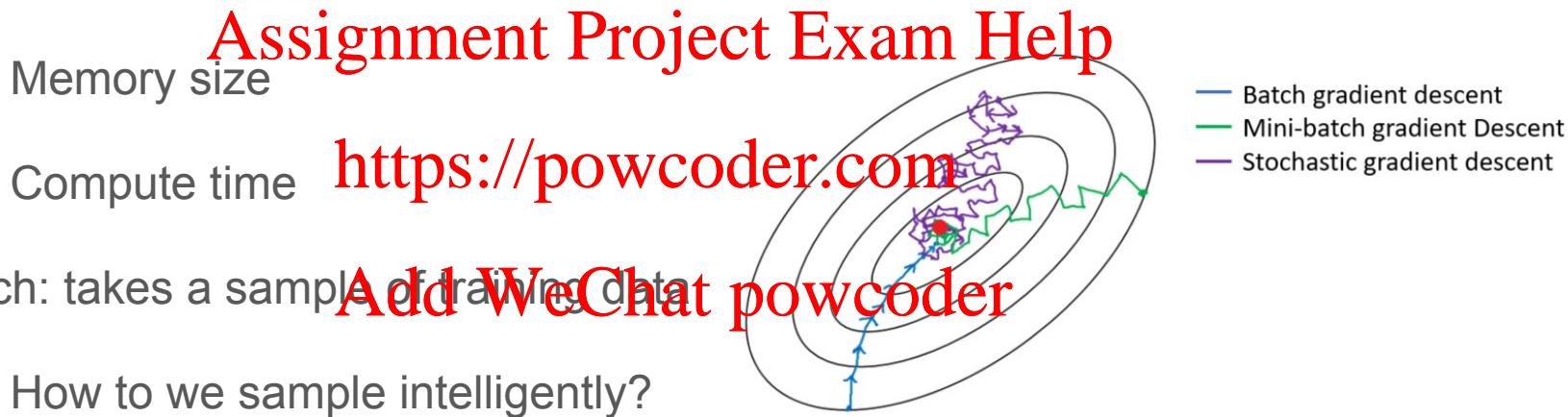
<https://powcoder.com>

Add WeChat powcoder



Mini-batch Gradient Descent

Expensive to compute gradient for large dataset



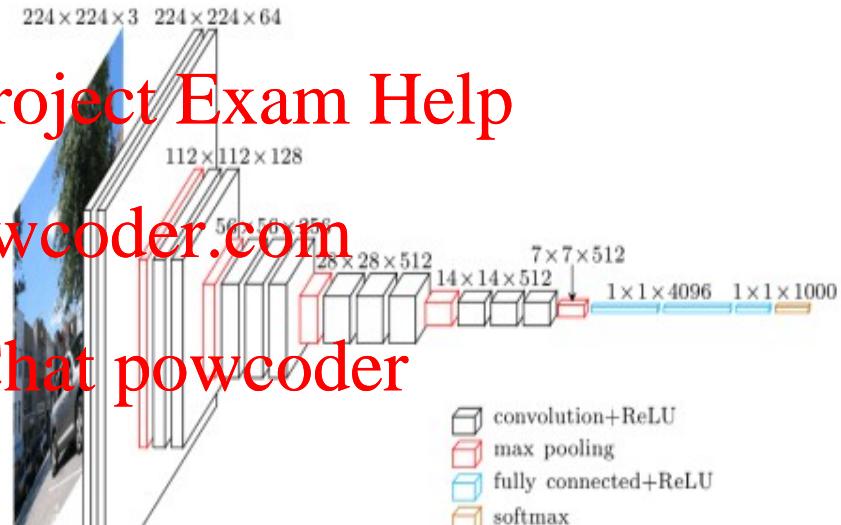
Is deeper better?

Deeper networks seem to be more powerful but harder to train

Assignment Project Exam Help

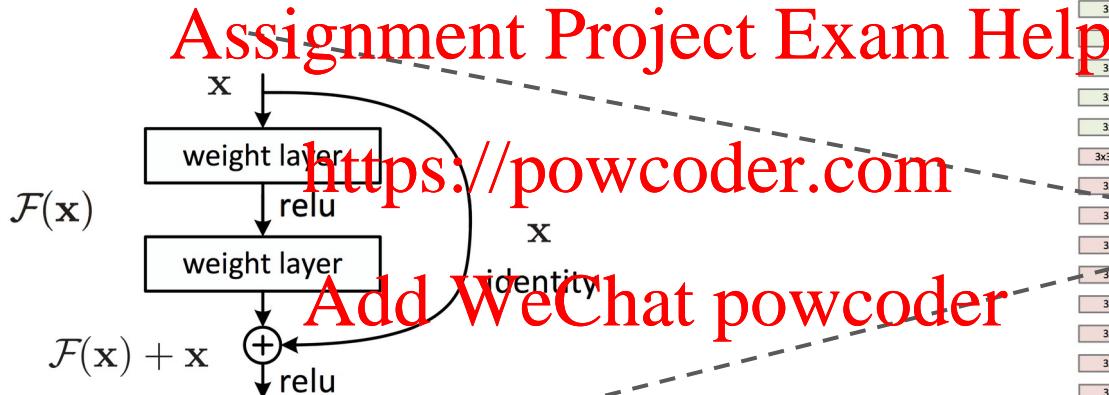
- Loss of information during forward propagation
- Loss of gradient info during back propagation

There are many ways to “keep the gradient going”



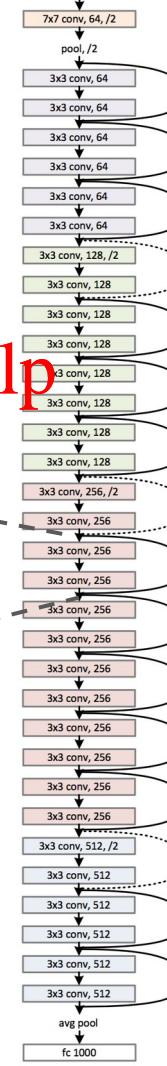
One Solution: skip connection

Connect the layers, create a gradient highway or information highway.



ResNet (2015)

Image credit: He et al. (2015)



Initialization

Can we initialize all neurons to zero?

If all the weights are same we will not be able to break symmetry of the network and all filters will end up learning the same thing.

Large numbers, might knock Relu units out.

Relu units once knocked out and their output is zero, their gradient flow also becomes zero.

We need small random numbers at initialization.

Variance : $1/\sqrt{n}$

Mean: 0

Popular initialization setups

(Xavier, Kaiming) (Uniform, Normal)

Dropout

What does cutting off some network connections do?

Assignment Project Exam Help

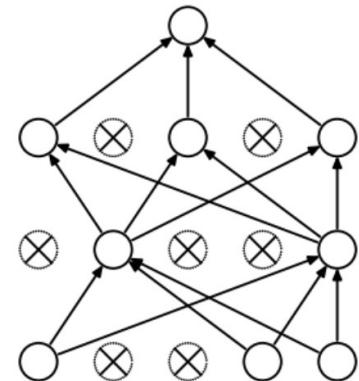
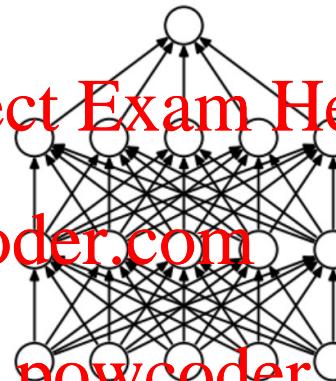
Trains multiple smaller networks in an ensemble.

<https://powcoder.com>

Can drop entire layer too!

Add WeChat powcoder

(a) Standard Neural Net



(b) After applying dropout.

Acts like a really good regularizer

More tricks for training

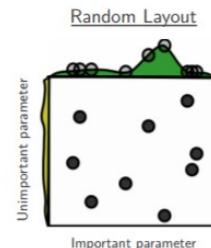
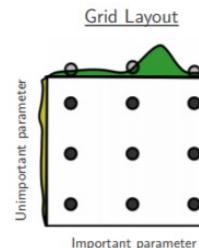
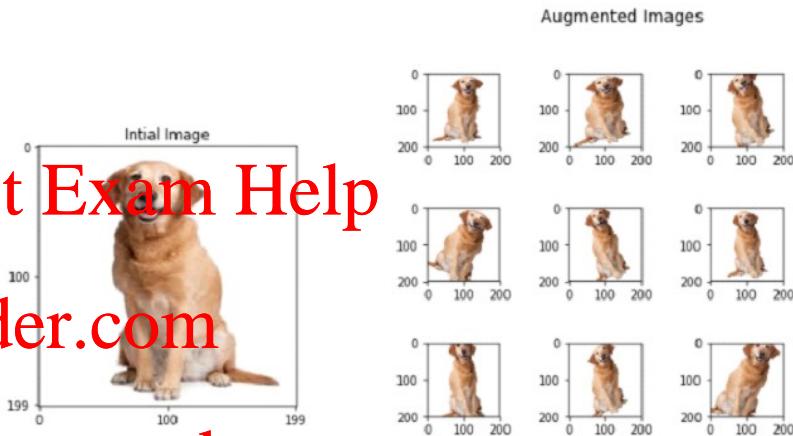
Data augmentation if your data set is smaller. This helps the network generalize more.

Assignment Project Exam Help

Early stopping if training loss goes above validation loss.

Add WeChat powcoder

Random hyperparameter search or grid search?



CNN sounds like fun!

What are some other areas of deep learning?

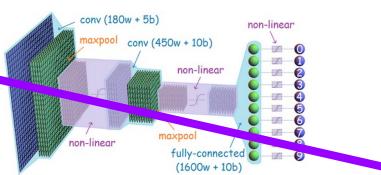
Recurrent NN
Sequential data

Assignment Project Exam Help

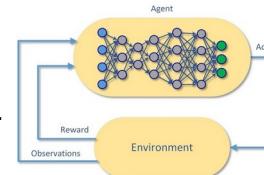
<https://powcoder.com>

Add WeChat powcoder

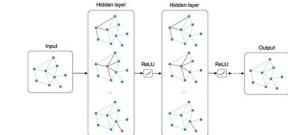
Convolutional NN



Deep RL



Graph NN



We can also have 1D architectures (remember this)

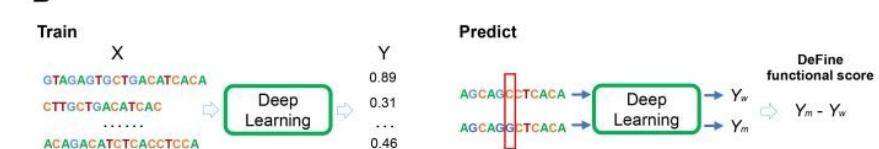
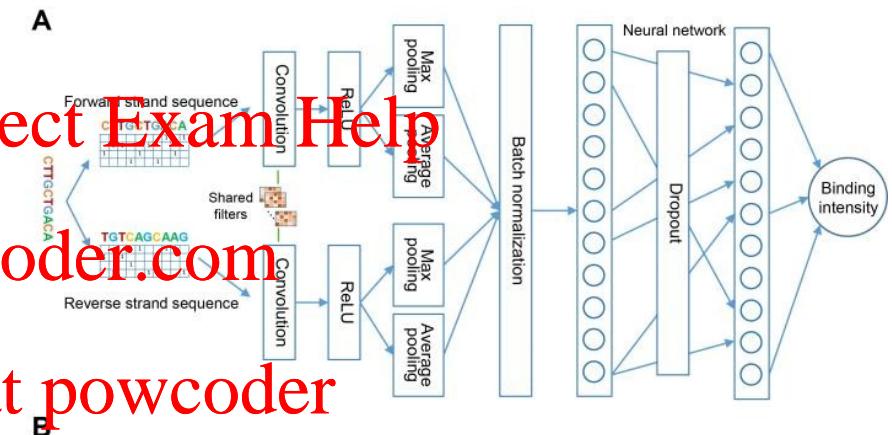
CNN works on any data where there is a local pattern

Assignment Project Exam Help

We use 1D convolutions on DNA sequences, text sequences, and music notes

Add WeChat powcoder

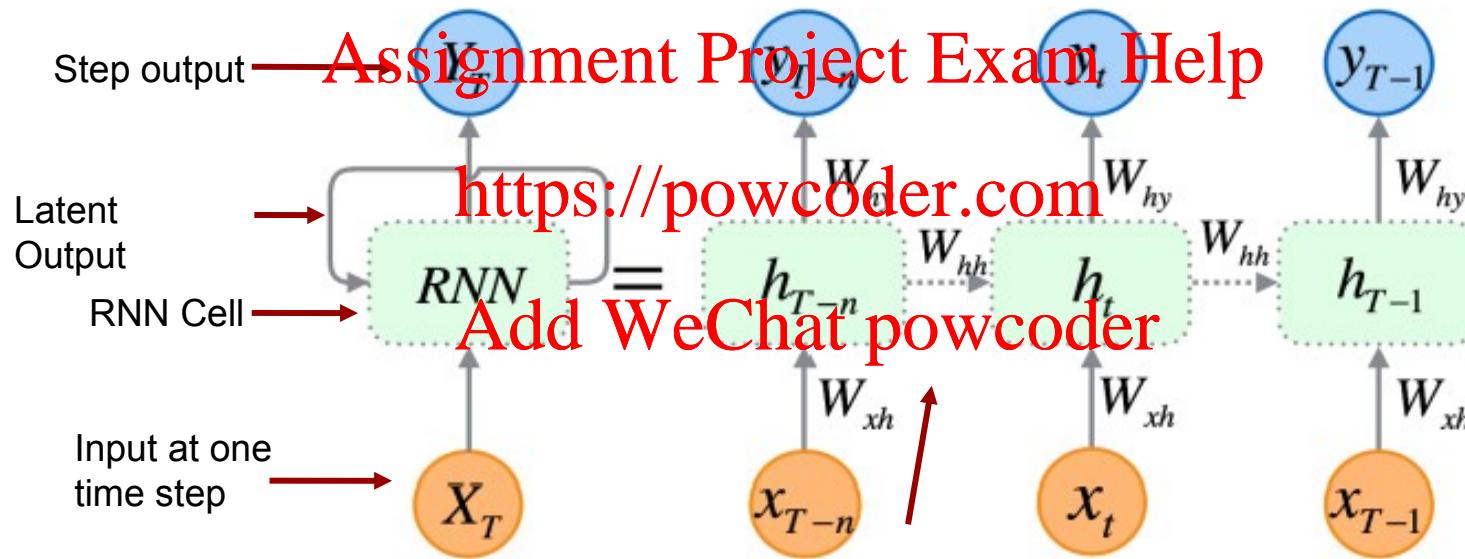
But what if time series has **causal dependency** or any kind of **sequential dependency**?



To address sequential dependency?

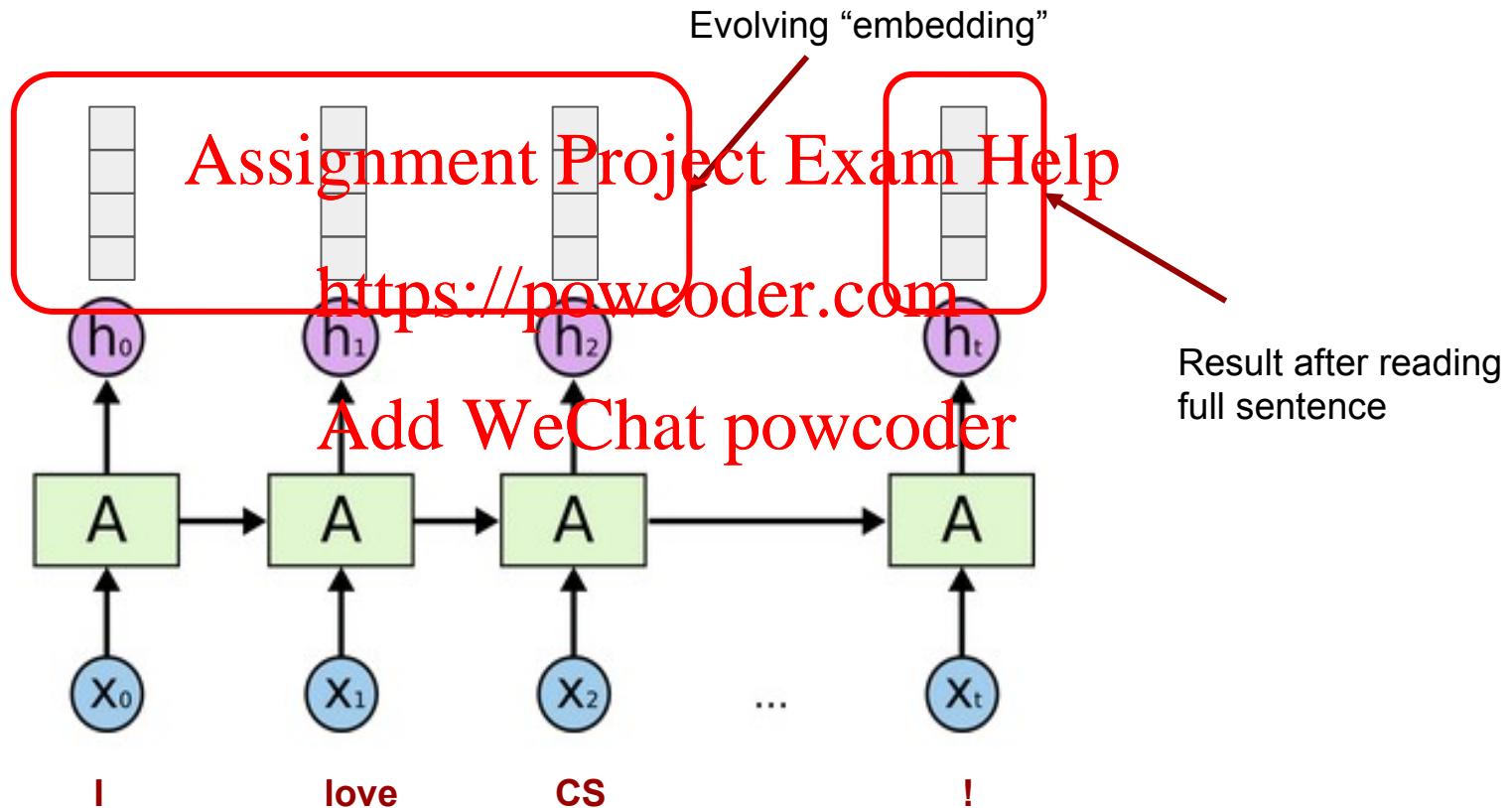
Use recurrent neural network (RNN)

Unrolling an RNN



The RNN Cell (Composed of W_{xh} and W_{hh} in this example) is really the same cell.
NOT many different cells like the filters of CNN.

How does RNN produce result?



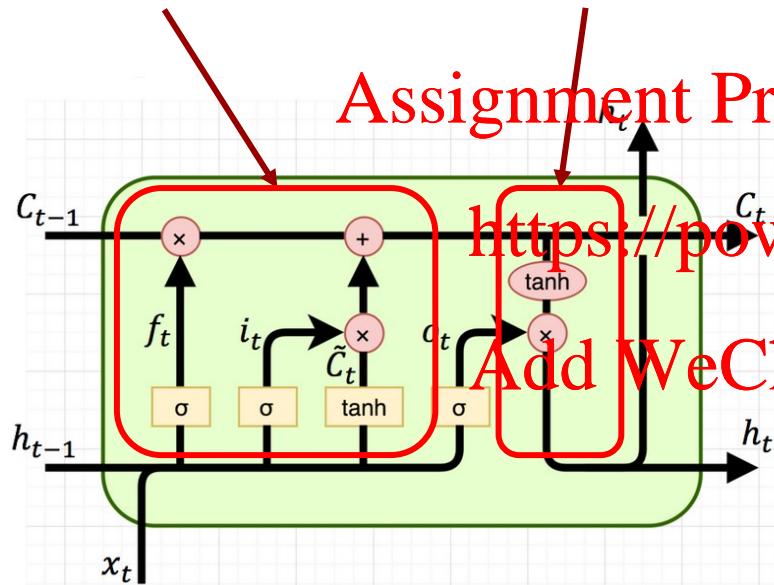
2 Typical RNN Cells

Store in “long term memory”

Response to current input

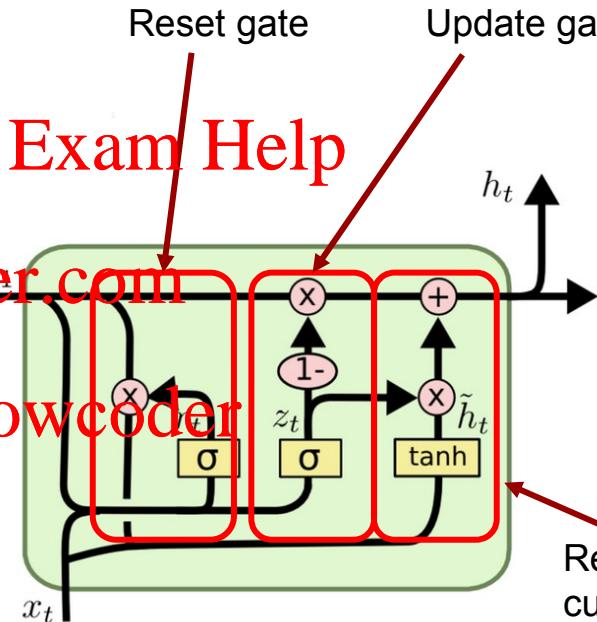
Reset gate

Update gate



Long Short Term Memory (**LSTM**)

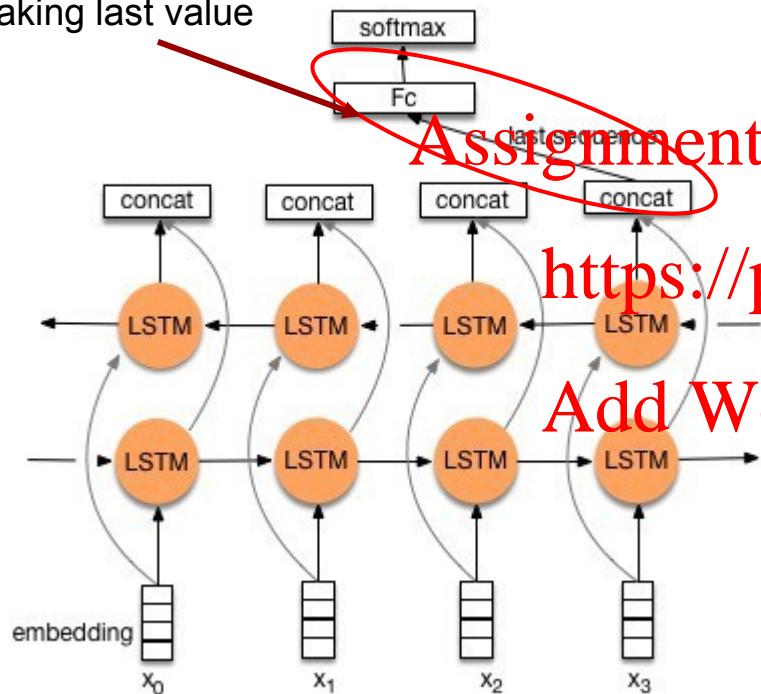
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



Gated Recurrent Unit (**GRU**)

Recurrent AND deep?

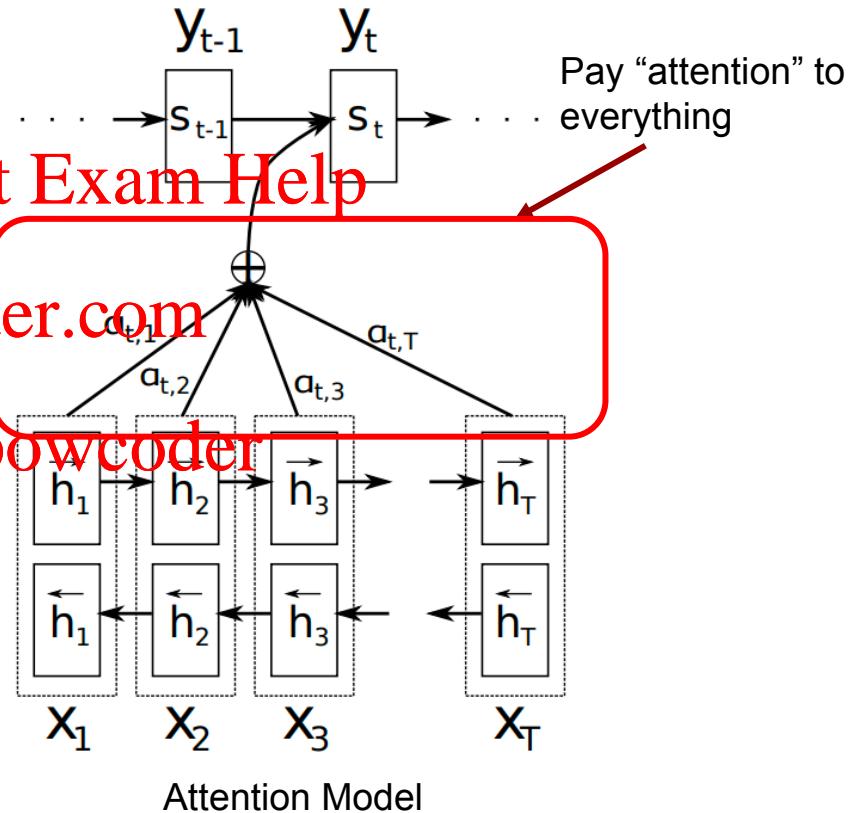
Taking last value



Assignment Project Exam Help

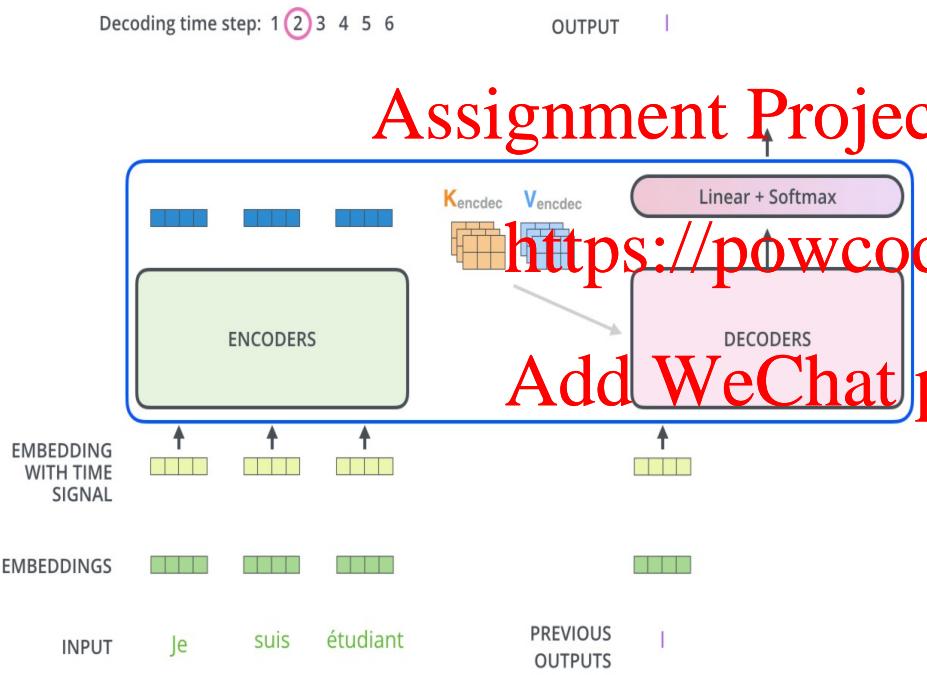
<https://powcoder.com>

Add WeChat powcoder



Transformer – Attention is All You Need!

Originally proposed for translation.



Assignment Project Exam Help

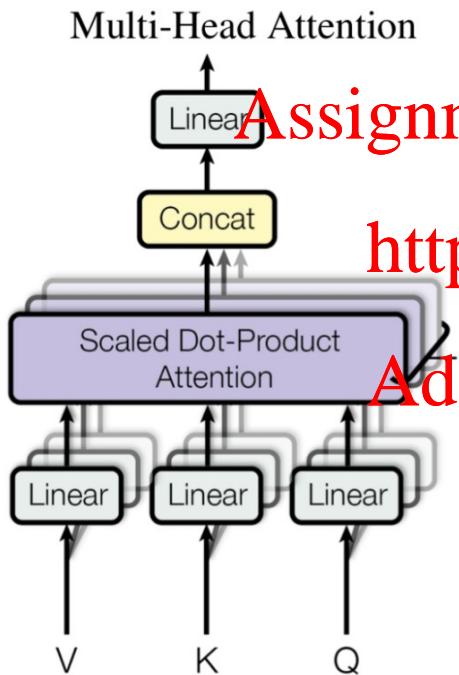
<https://powcoder.com>

Add WeChat **powcoder**

Encoder computes hidden representations for each word in the input sentence
Applies **self attention**.

Decoder makes sequential prediction similar as in RNN
At each time step, it predicts the next word based on its previous predictions (partial sentence).
Applies **self attention** and **attention on encoder outputs**.

Transformer – Attention is All You Need!



The dot product in softmax below computes how each word of sequence 1 (Q) is influenced by all the other words in the sequence 2 (K).

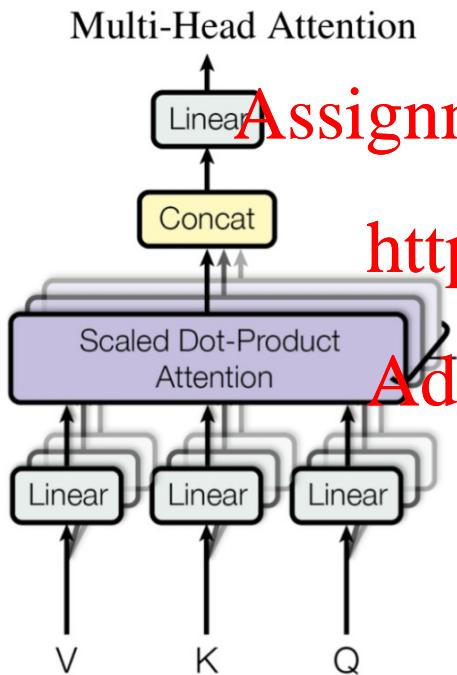
<https://powcoder.com>

Add WeChat powcoder

Considering the different importance, we computed a weighted sum of the information in the sequence 2 (V) to use in computing the hidden representation of sequence 1.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Transformer – Attention is All You Need!



The dot product in softmax below computes how each word of sequence 1 (Q) is influenced by all the other words in the sequence 2 (K).

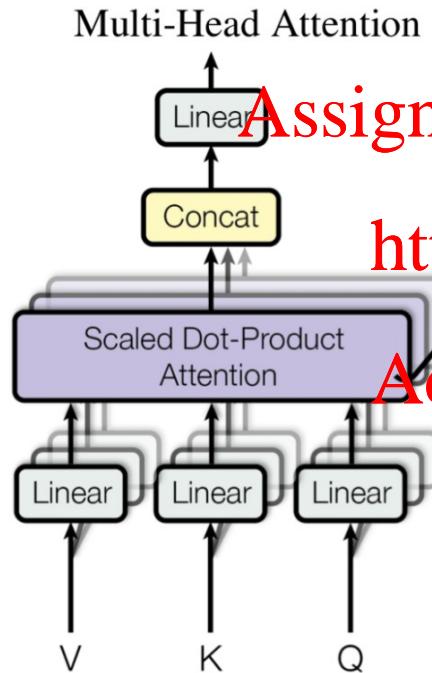
<https://powcoder.com>

Add WeChat powcoder

Considering the different importance, we computed a weighted sum of the information in the sequence 2 (V) to use in computing the hidden representation of sequence 1.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Transformer – Attention is All You Need!



Assignment Project Exam Help

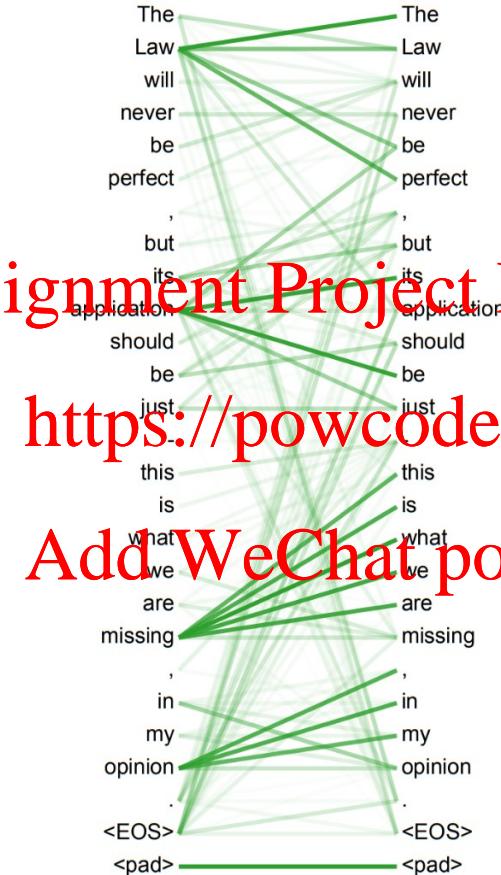
Multiple heads!

-- Similar as how you have multiple filters in CNN

Add WeChat [powcoder](https://powcoder.com)

Loss of sequential order?

Positional encoding! (often use sine waves)



Examples of attention scores from two different self-attention heads.

Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder

References:

<https://arxiv.org/pdf/1706.03762.pdf>
<https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04>
<https://towardsdatascience.com/transformers-141e32e69591>

<https://towardsdatascience.com/transformers-explained-visually-part-2-how-it-works-step-by-step-b49fa4a64f34>

SOTA Example – GPT3

A "whatpu" is a small, furry animal native to Tanzania. An example of a sentence that uses the word whatpu is:

We were traveling in Africa and we saw these very cute whatpus.

To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses the word farduddle is:

One day when I was playing tag with my little sister, she got really excited and she started doing these crazy farduckles.

A "yalubalu" is a type of vegetable that looks like a big pumpkin. An example of a sentence that uses the word yalubalu is:

I was on a trip to Africa and I tried this yalubalu vegetable that was grown in a garden there. It was delicious.

A "Burringo" is a car with very fast acceleration. An example of a sentence that uses the word Burringo is:

In our garage we have a Burringo that my father drives to work every day.

A "Gigamuru" is a type of Japanese musical instrument. An example of a sentence that uses the word Gigamuru is:

I have a Gigamuru that my uncle gave me as a gift. I love to play it at home.

To "screeg" something is to swing a sword at it. An example of a sentence that uses the word screeg is:

We screeghed at each other for several minutes and then we went outside and ate ice cream.

Add WeChat **powcoder**

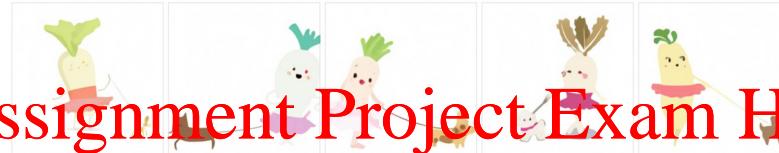
<https://powcoder.com>

SOTA Example – DALLE

TEXT PROMPT

an illustration of a baby daikon radish in a tutu walking a dog

AI-GENERATED IMAGES



Edit prompt or view more images ↴

Assignment Project Exam Help

TEXT PROMPT

an armchair in the shape of an avocado

AI-GENERATED IMAGES



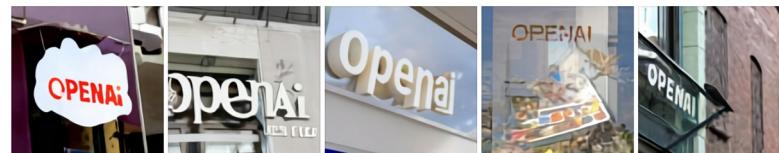
Edit prompt or view more images ↴

Add WeChat powcoder

TEXT PROMPT

a store front that has the word 'openai' written on it [...]

AI-GENERATED IMAGES



Edit prompt or view more images ↴

SOTA Example – GPT3

Describe a layout.

Just describe any layout you want, and it'll try to render below!

a button that looks like a watermelon

Generate

```
<button style={{backgroundColor: 'pink', border: '2px solid green', borderRadius: '50%', padding: 20, width: 100, height: 100}}>Watermelon</button>
```



Add WeChat powcoder

a button for every color of the rainbow

Generate

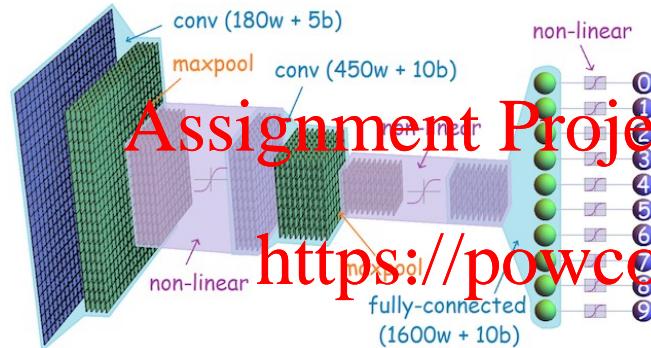
```
<div style={{backgroundColor: 'red', padding: 20}}>Red</div><div style={{backgroundColor: 'orange', padding: 20}}>Orange</div><div style={{backgroundColor: 'yellow', padding: 20}}>Yellow</div><div style={{backgroundColor: 'green', padding: 20}}>Green</div><div style={{backgroundColor: 'blue', padding: 20}}>Blue</div><div style={{backgroundColor: 'indigo', padding: 20}}>Indigo</div>
```



https://powcoder.com

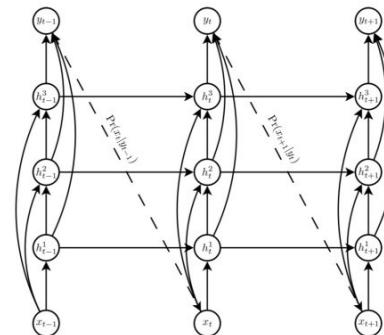
More? Take CS230, CS236, CS231N, CS224N

Convolutional NN
Image



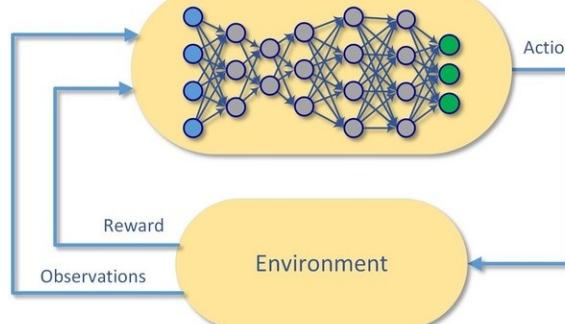
Assignment Project Exam Help
<https://powcoder.com>

Recurrent NN
Time Series

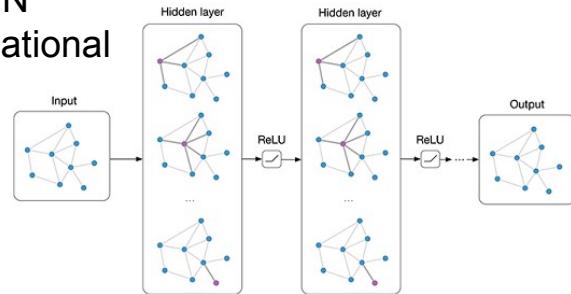


Add WeChat powcoder

Deep RL
Control System

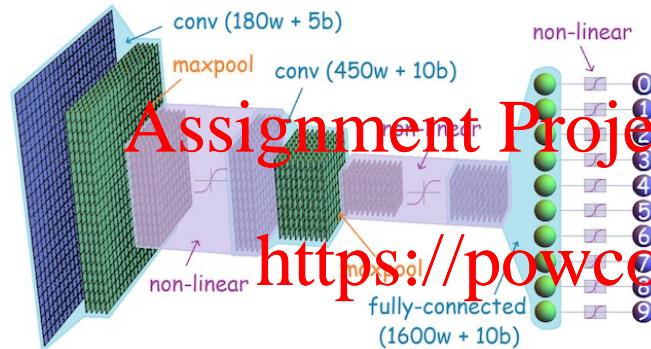


Graph NN
Networks/Relational



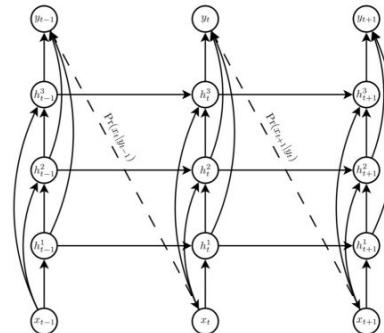
Not today, but take CS234 and CS224W

Convolutional NN
Image



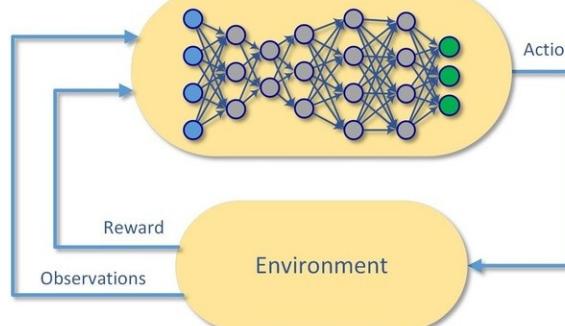
Assignment Project Exam Help
<https://powcoder.com>

Recurrent NN
Time Series

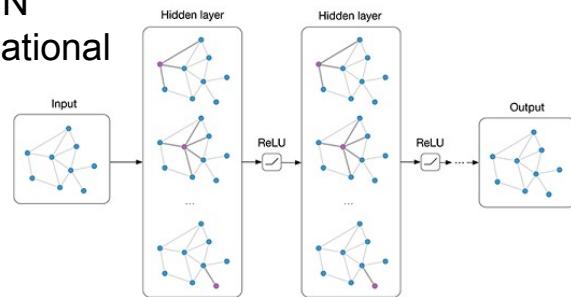


Add WeChat powcoder

Deep RL
Control System



Graph NN
Networks/Relational



Tools for deep learning



Keras

Assignment Project Exam Help

+theano

Add WeChat powcoder

P Y T H O N

Popular Tools

Specialized Groups



m x net

Microsoft

C N T K

\$50 not enough! Where can I get free stuff?

Google Colab

Free (limited GPU access)

Assignment Project Exam Help

Works nicely with Tensorflow

<https://powcoder.com>

Links to Google Drive

Register a new Google Cloud account

Add WeChat ~~to SAVE money~~ powcoder

=> Instant \$300??

CLOSE your GPU instance

=> AWS free tier (limited compute)

=> Azure education account, \$200?

~\$1 an hour

Good luck!

Well, have fun too! <https://powcoder.com>

Assignment Project Exam Help

Add WeChat powcoder