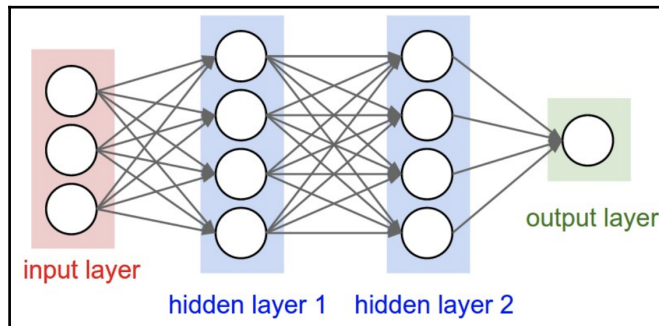
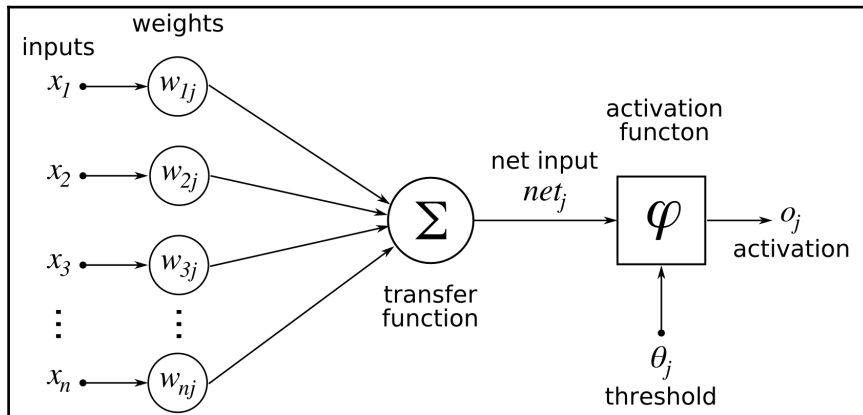
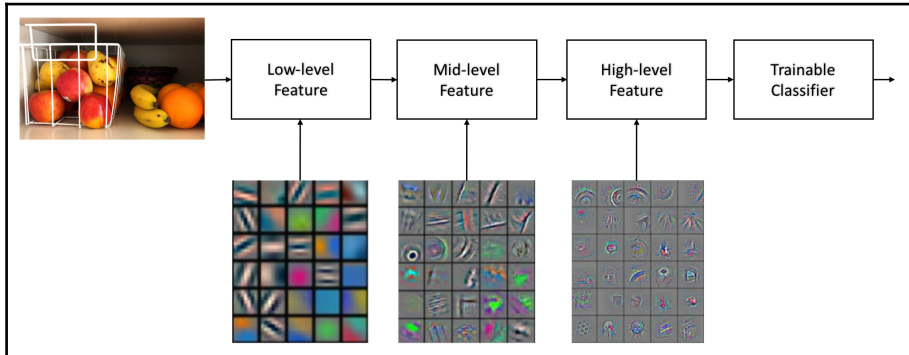
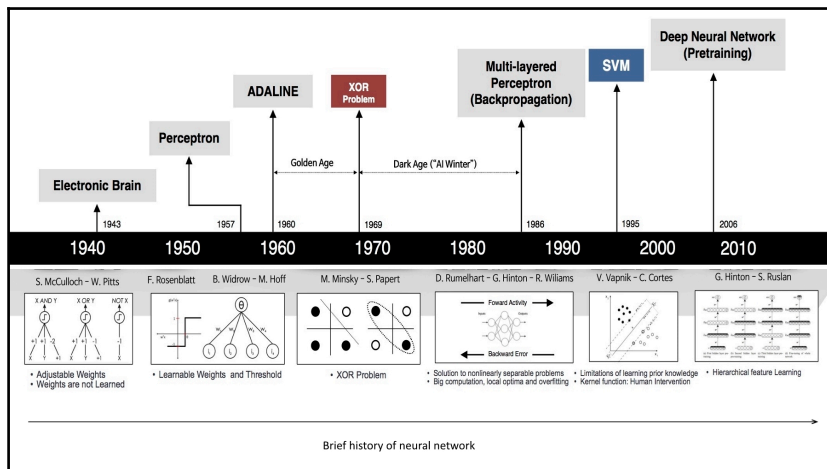
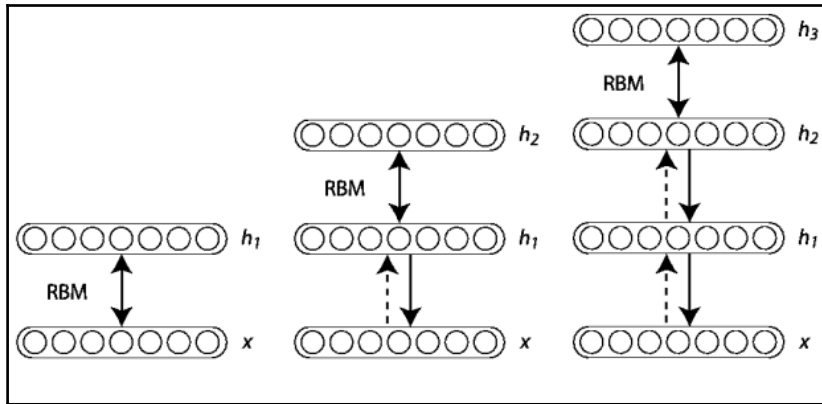
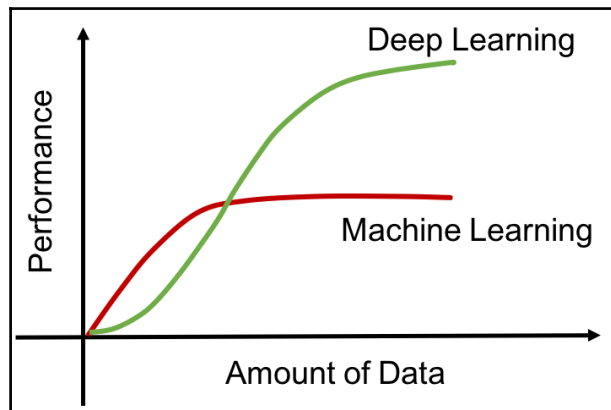
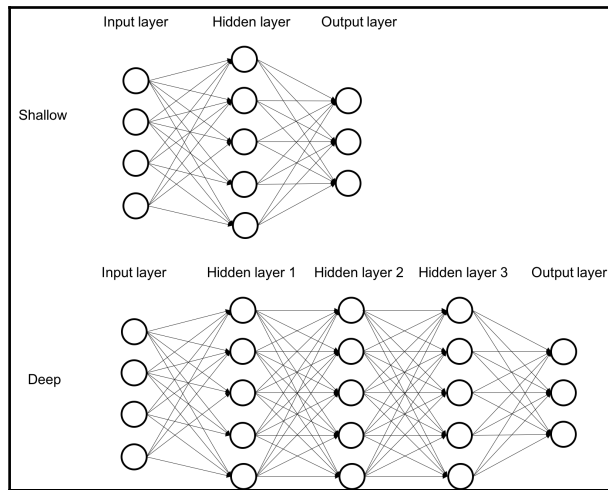
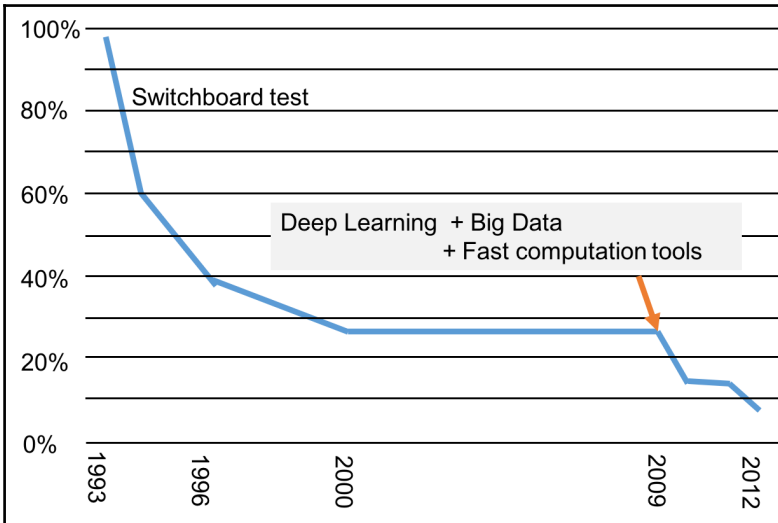
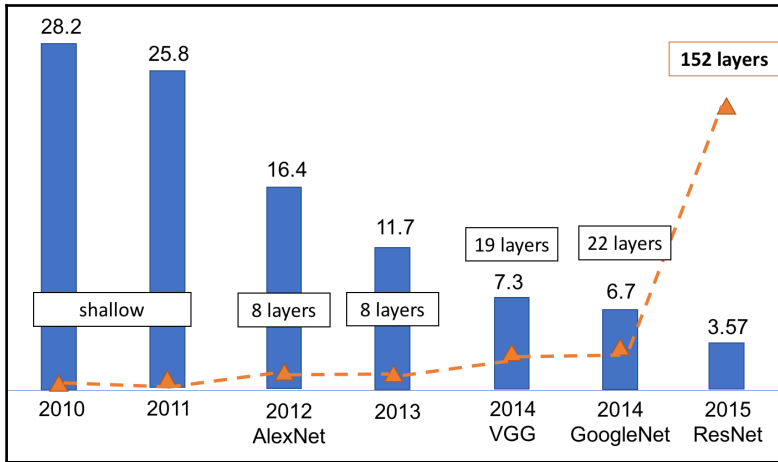


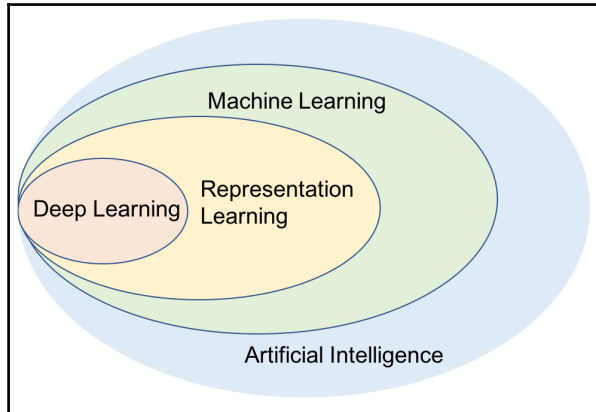
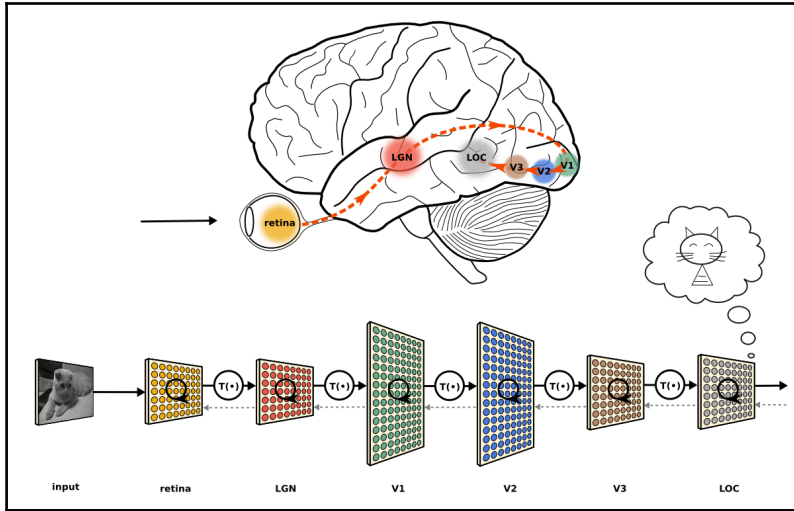
Chapter 1: Why Deep Learning?













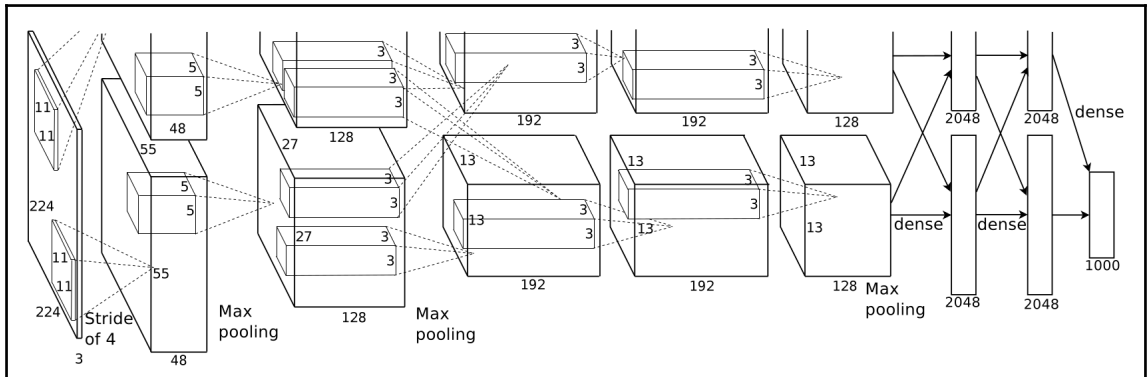
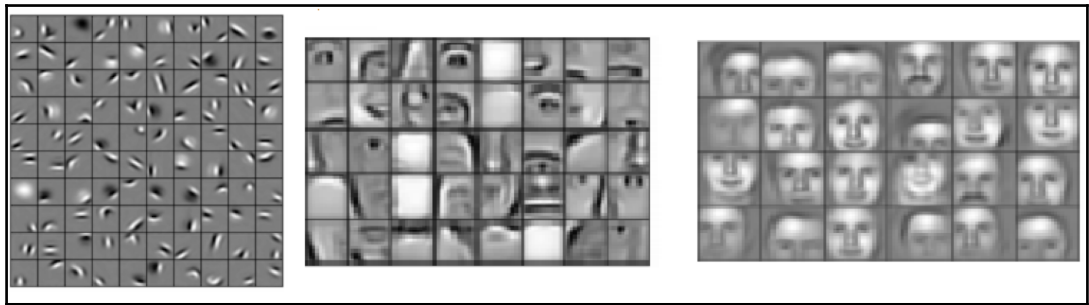
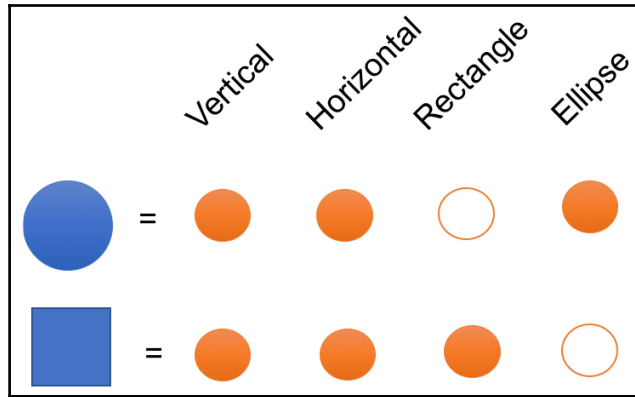


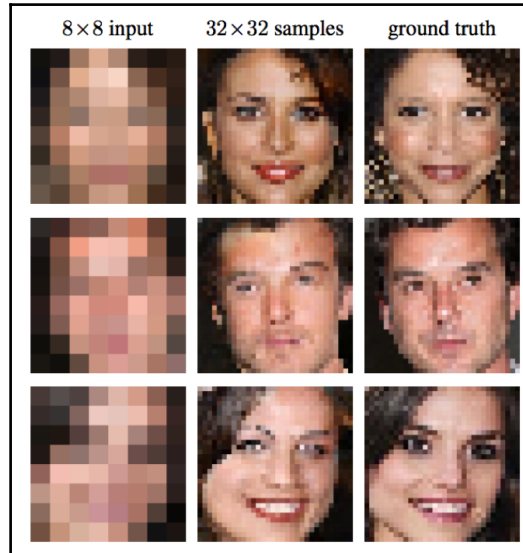
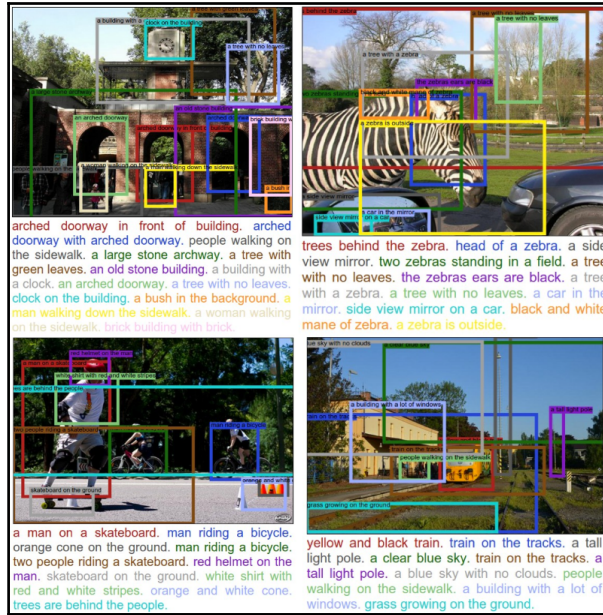


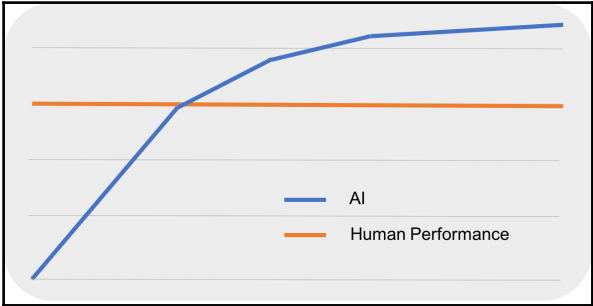
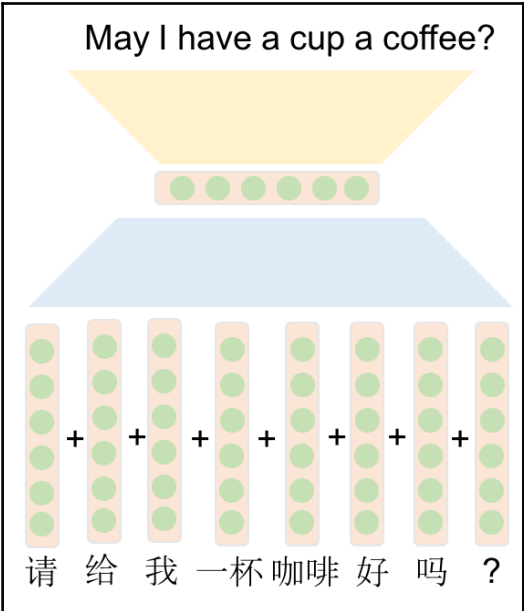


	Vertical Rectangle	Horizontal Rectangle	Vertical Ellipse	Horizontal Ellipse
No Pattern	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

	Vertical	Horizontal	Rectangle	Ellipse
No Pattern	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>



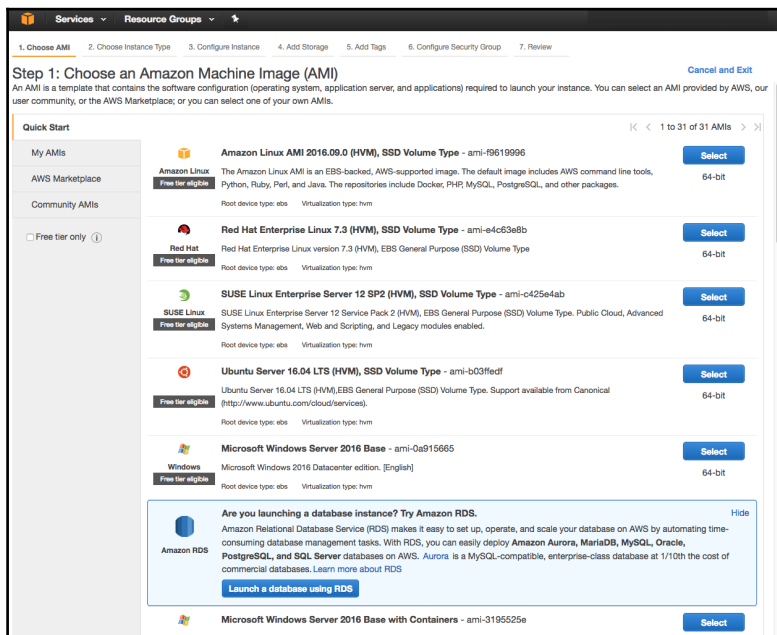




Chapter 2: Getting Yourself Ready for Deep Learning

	High	Medium	Low
Data parallelism	Almost entire method is data parallel* (75–100%).	More than half of the method is data parallel (50–75%).	None or up to half of the method is data parallel (0–50%).
Thread count	The thread count is equal to or more than the number of pixels/voxels in the image.	The thread count is in the thousands.	The thread count is less than a thousand.
Branch divergence	More than 10% of the AUEs** have branch divergence and the code complexity in the branch is substantial.	Less than 10% of the AUEs have branch divergence, but the code complexity is substantial.	The code complexity in the branches is low.
Memory usage	More than 5N***	From 2N to 5N.	2N or less.
Synchronization	Global synchronization is performed more than hundred times. This is usually true for iterative methods.	Global synchronization is performed between 10 and 100 times.	Only a few global or local synchronizations.

* Data Parallel : An algorithm that can perform the same instructions on multiple data elements in parallel is said to be data parallel.
 ** AUE: An AUE is thus a group of threads that are all executed atomically on thread processors in the same core. Nvidia calls them wraps while AMD calls them wavefronts.
 *** N is the total number of pixels/voxels in the image



Services Resource Groups

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Add Tags 6. Configure Security Group 7. Review

Step 2: Choose an Instance Type

<input type="checkbox"/>	Compute optimized	c3.large	2	3.75	2 x 16 (SSD)	-	Moderate
<input type="checkbox"/>	Compute optimized	c3.xlarge	4	7.5	2 x 40 (SSD)	Yes	Moderate
<input type="checkbox"/>	Compute optimized	c3.2xlarge	8	15	2 x 80 (SSD)	Yes	High
<input type="checkbox"/>	Compute optimized	c3.4xlarge	16	30	2 x 160 (SSD)	Yes	High
<input type="checkbox"/>	Compute optimized	c3.8xlarge	32	60	2 x 320 (SSD)	-	10 Gigabit
<input checked="" type="checkbox"/>	GPU instances	g2.2xlarge	8	15	1 x 60 (SSD)	Yes	High
<input type="checkbox"/>	GPU instances	g2.8xlarge	32	60	2 x 120 (SSD)	-	10 Gigabit
<input type="checkbox"/>	Memory optimized	r4.large	2	15.25	EBS only	Yes	High
<input type="checkbox"/>	Memory optimized	r4.xlarge	4	30.5	EBS only	Yes	High
<input type="checkbox"/>	Memory optimized	r4.2xlarge	8	61	EBS only	Yes	High
<input type="checkbox"/>	Memory optimized	r4.4xlarge	16	122	EBS only	Yes	High
<input type="checkbox"/>	Memory optimized	r4.8xlarge	32	244	EBS only	Yes	10 Gigabit
<input type="checkbox"/>	Memory optimized	r4.16xlarge	64	488	EBS only	Yes	20 Gigabit
<input type="checkbox"/>	Memory optimized	r3.large	2	15	1 x 32 (SSD)	-	Moderate
<input type="checkbox"/>	Memory optimized	r3.xlarge	4	30.5	1 x 80 (SSD)	Yes	Moderate
<input type="checkbox"/>	Memory optimized	r3.2xlarge	8	61	1 x 160 (SSD)	Yes	High
<input type="checkbox"/>	Memory optimized	r3.4xlarge	16	122	1 x 320 (SSD)	Yes	High
<input type="checkbox"/>	Memory optimized	r3.8xlarge	32	244	2 x 320 (SSD)	-	10 Gigabit
<input type="checkbox"/>	Memory optimized	x1.16xlarge	64	976	1 x 1920	Yes	10 Gigabit
<input type="checkbox"/>	Memory optimized	x1.32xlarge	128	1952	2 x 1920	Yes	20 Gigabit
<input type="checkbox"/>	Storage optimized	d2.xlarge	4	30.5	3 x 2048	Yes	Moderate
<input type="checkbox"/>	Storage optimized	d2.2xlarge	8	61	6 x 2048	Yes	High
<input type="checkbox"/>	Storage optimized	d2.4xlarge	16	122	12 x 2048	Yes	High

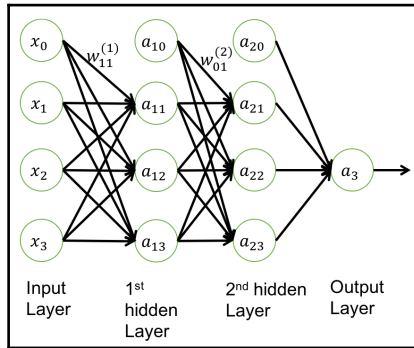
Cancel Previous **Review and Launch** Next: Configure Instance Details

```

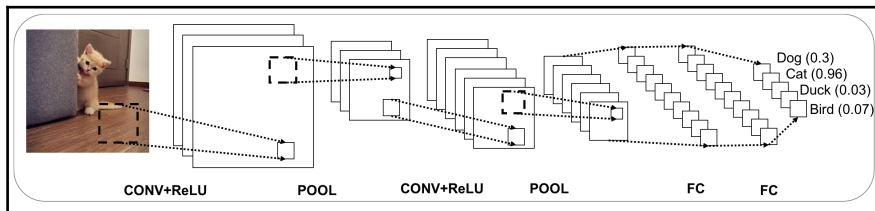
Last login: Tue Jan 16 17:20:00 on ttys001
Anurag:~ anuragbhardwaj$ python tensor-toy.py
dyld: warning, LC_RPATH $ORIGIN/../../_solib_darwin_x86_64/_U_S_Stensorflow_Spython_C_Upywrap_Utensorflow_Uinternal.so___Utensorflow in /Library/Python/2.7/site-packages/tensorflow/python/_pywrap_tensorflow_internal.so being ignored in restricted program because it is a relative path
Couldn't import dot_parser, loading of dot files will not be possible.
2018-01-16 17:39:57.587007: I tensorflow/core/platform/cpu_feature_guard.cc:137]
Your CPU supports instructions that this TensorFlow binary was not compiled to use: SSE4.2 AVX AVX2 FMA
('Value after running graph:', 10.0)
Anurag:~ anuragbhardwaj$

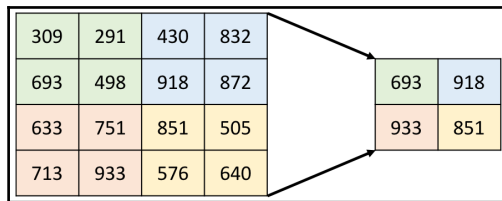
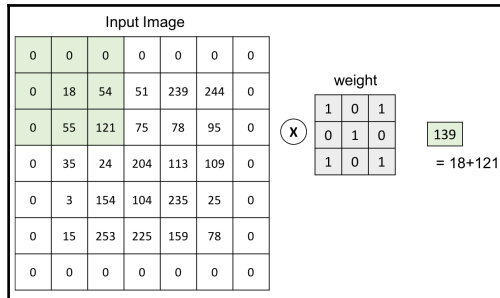
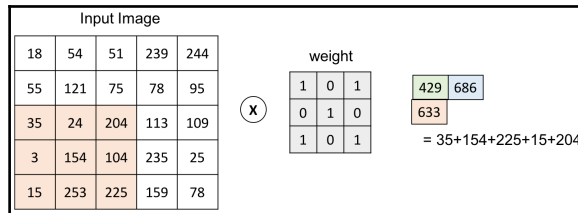
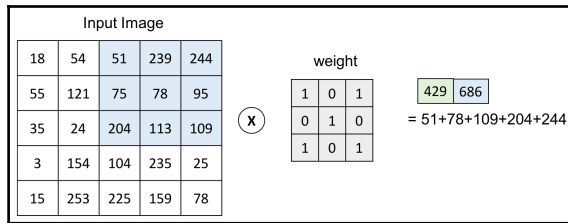
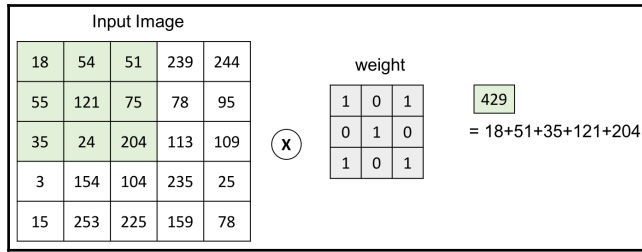
```

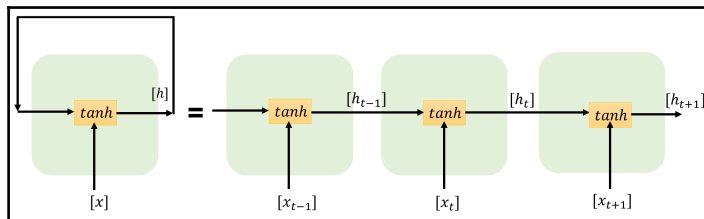
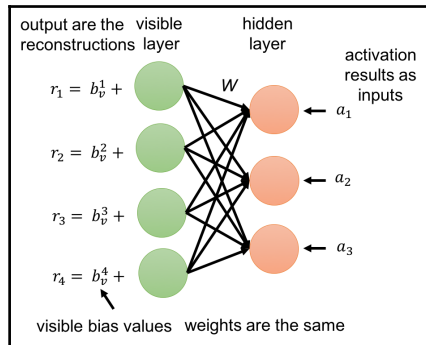
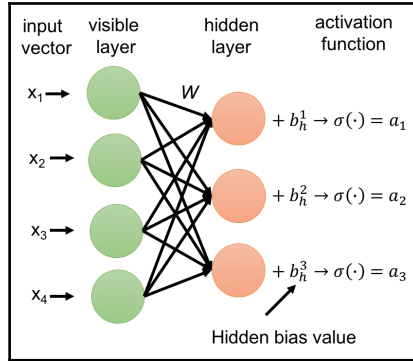
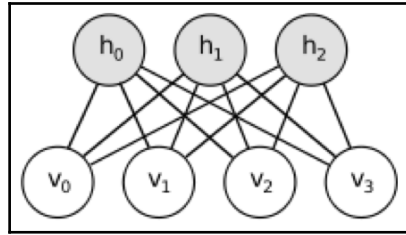
Chapter 3: Getting Started with Neural Networks

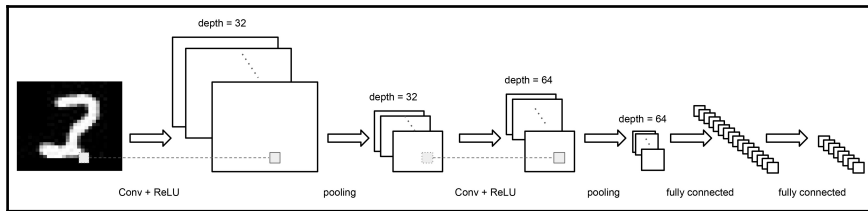
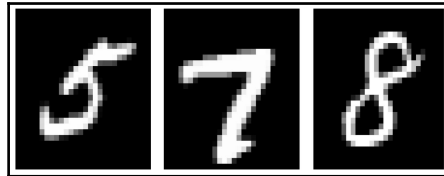
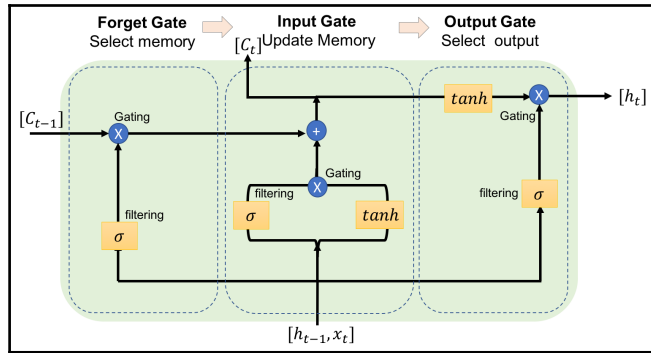


Name	Equation	Derivative	1-D Graph	1-D Graph(derivative)
Binary Step	$\sigma(x) = \begin{cases} 1, & x > 0 \\ 0.5, & x = 0 \\ 0, & x < 0 \end{cases}$	$\sigma'(x) = \begin{cases} 0, & x \neq 0 \\ ?, & x = 0 \end{cases}$		
Identity	$\sigma(x) = x$	$\sigma'(x) = 1$		
Sigmoid	$\sigma(x) = \frac{1}{1 + e^{-x}}$	$\sigma'(x) = \sigma(x)(1 - \sigma(x))$		
Tanh	$\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$\sigma'(x) = 1 - \sigma(x)^2$		
Rectified Linear (ReLU)	$\sigma(x) = \max(0, x)$	$\sigma'(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$		
Leaky ReLU	$\sigma(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases}$	$\sigma'(x) = \begin{cases} 1, & x \geq 0 \\ \alpha, & x < 0 \end{cases}$		

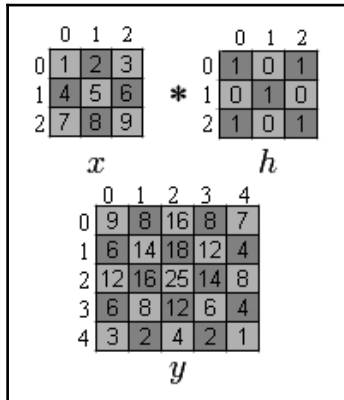
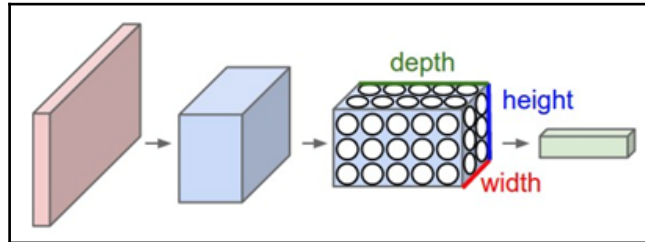
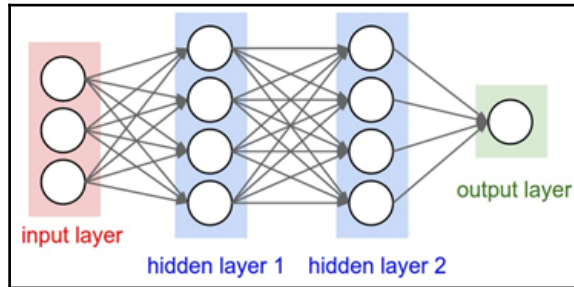


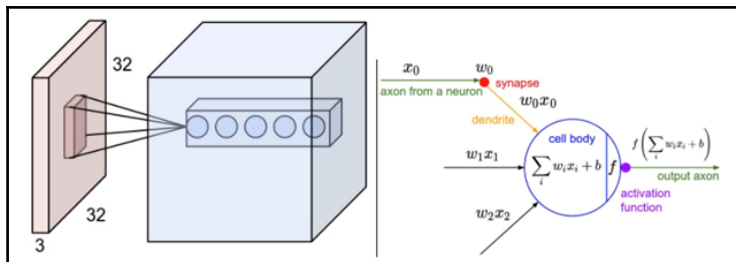
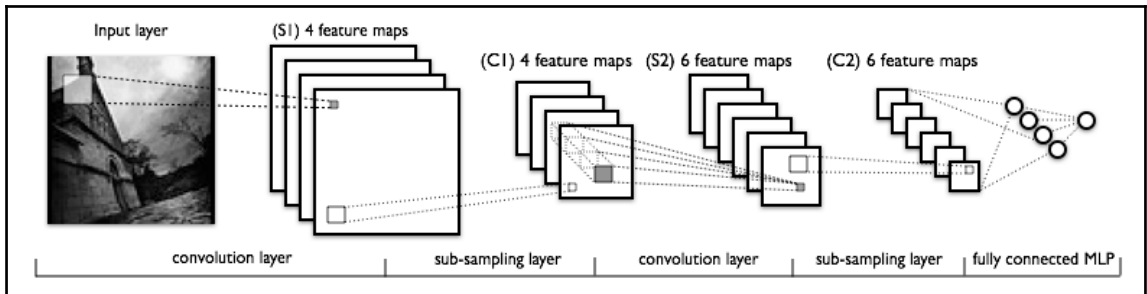
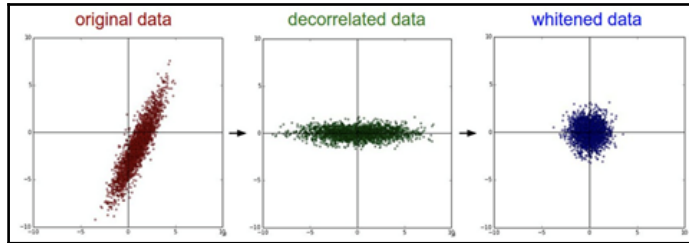
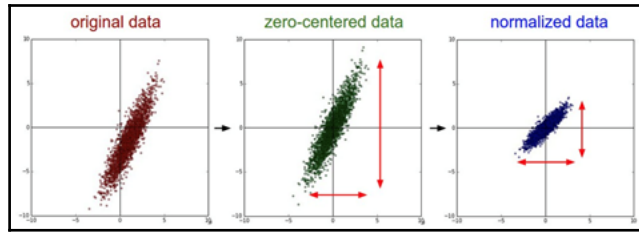


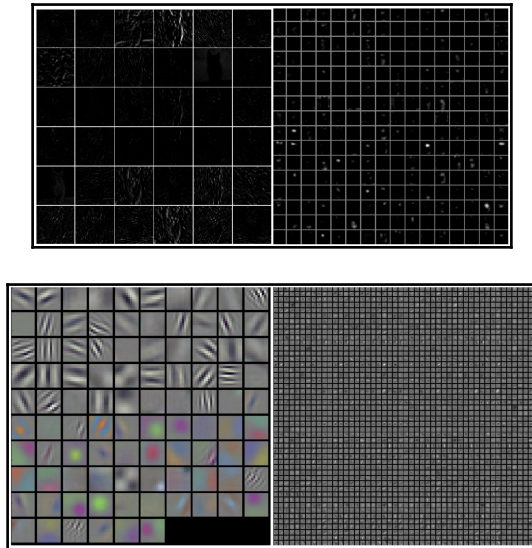
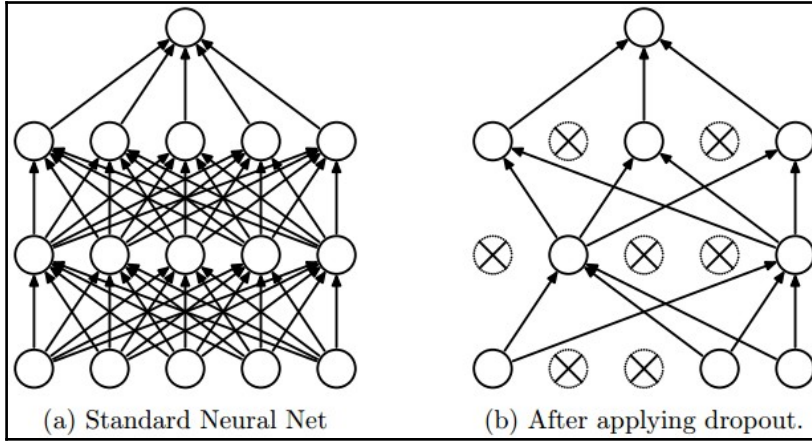
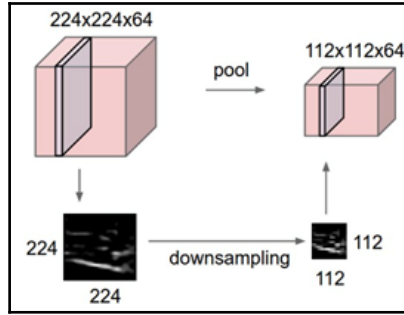


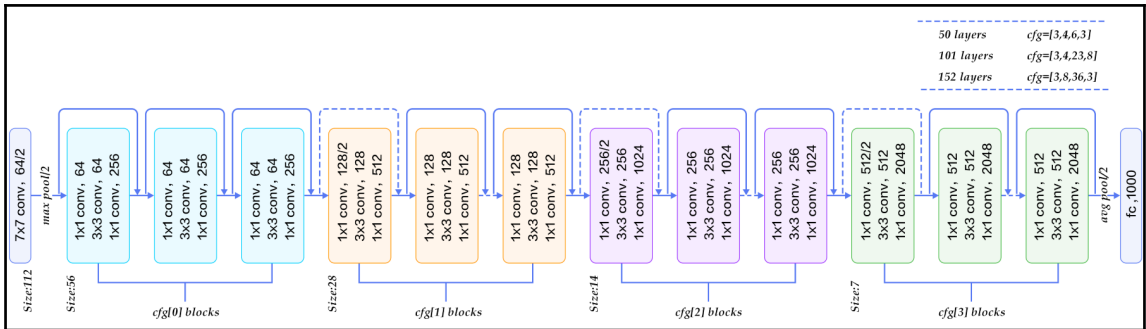
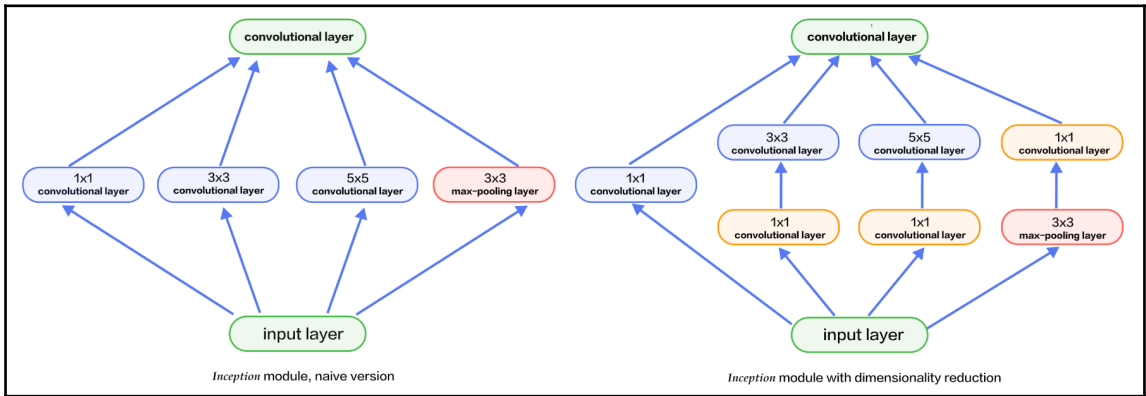
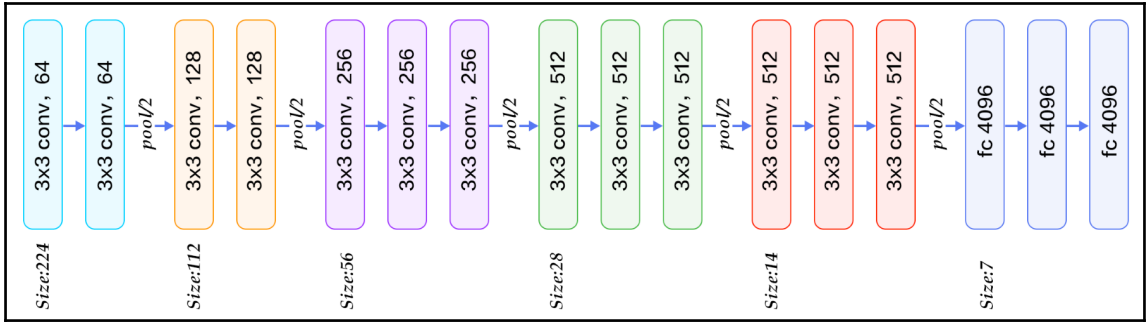


Chapter 4: Deep Learning in Computer Vision

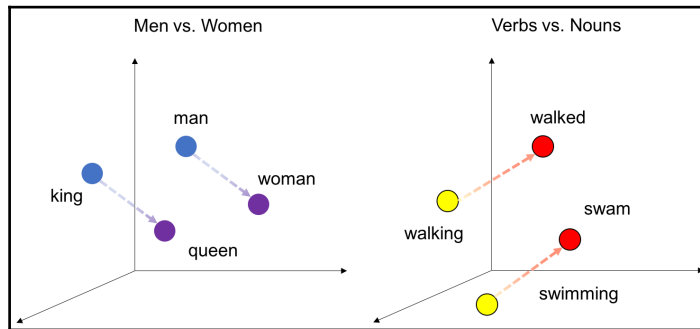
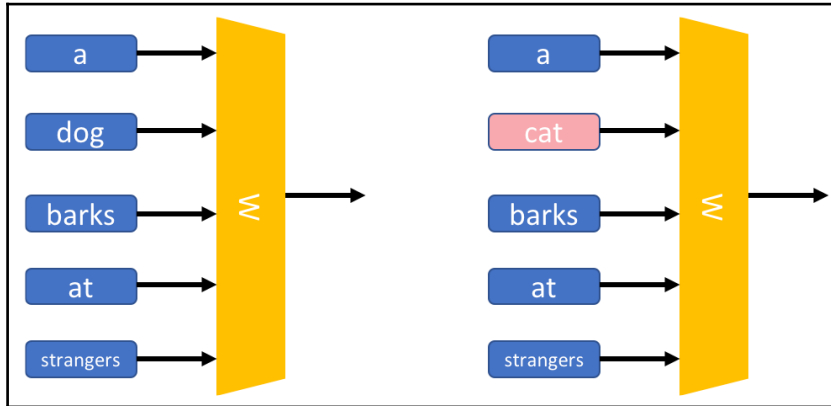




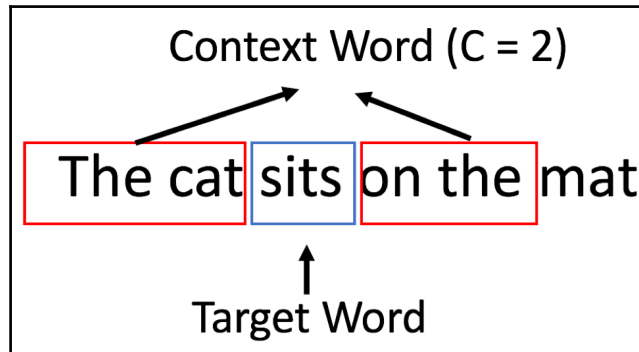
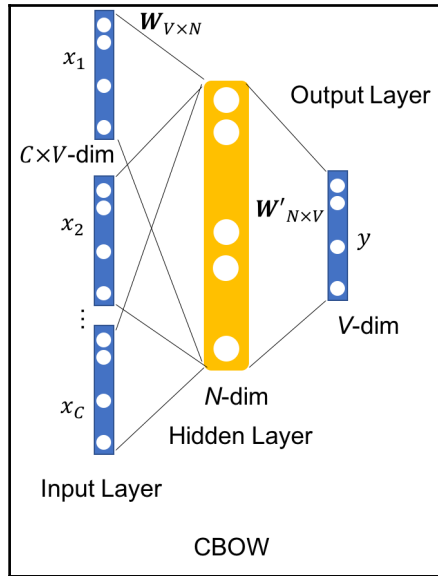




Chapter 5: NLP - Vector Representation



Type of relationship	Word Pair 1		Word Pair 2	
Common capital city	Athens	Greece	Oslo	Norway
All capital cities	Astana	Kazakhstan	Harare	Zimbabwe
Currency	Angola	kwanza	Iran	rial
City-in-state	Chicago	Illinois	Stockton	California
Man-Woman	brother	sister	grandson	granddaughter
Adjective to adverb	apparent	apparently	rapid	rapidly
Opposite	possibly	impossibly	ethical	unethical
Comparative	great	greater	tough	tougher
Superlative	easy	easiest	lucky	luckiest
Present Participle	think	thinking	read	reading
Nationality adjective	Switzerland	Swiss	Cambodia	Cambodian
Past tense	walking	walked	swimming	swam
Plural nouns	mouse	mice	dollar	dollars
Plural verbs	work	works	speak	speaks



The quick brown fox jumps over the lazy dog.

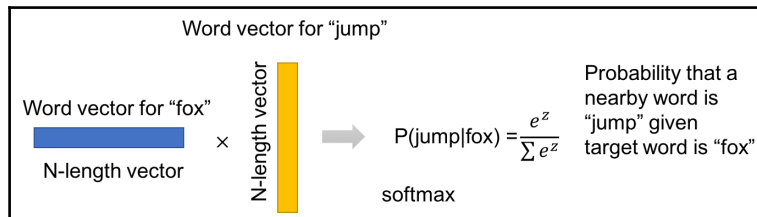
Target	Context	
quick	the	} Training Data (Positive)
quick	brown	
quick	fox	

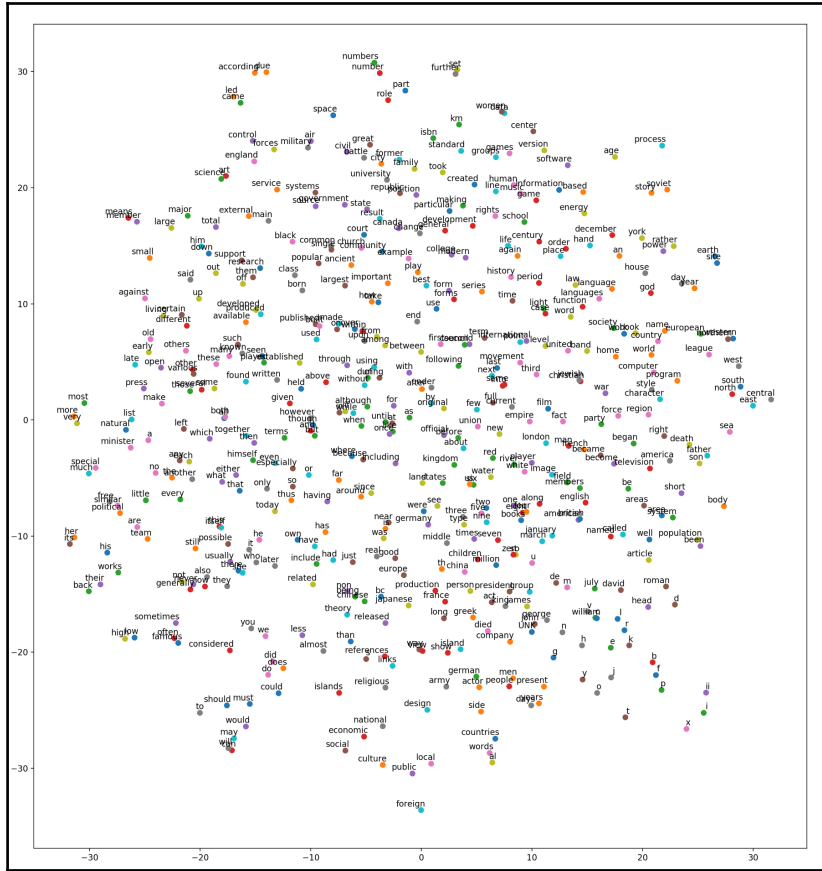
The quick brown fox jumps
over the lazy dog.

Target	Context	
fox	quick	} Training Data (Positive)
fox	brown	
fox	jumps	
fox	over	

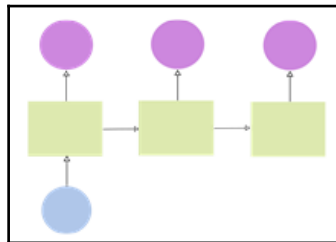
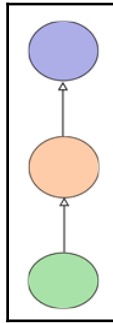
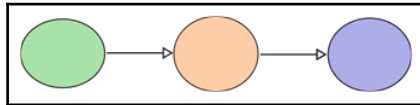
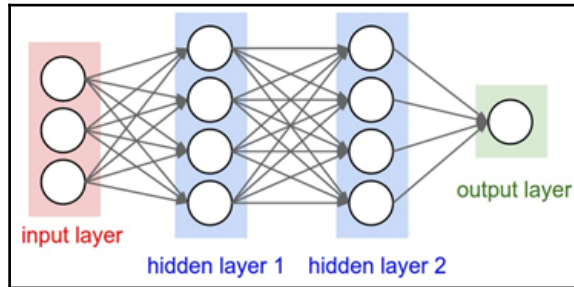
The quick brown fox jumps
over the lazy dog.

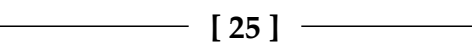
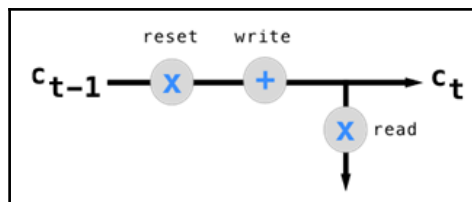
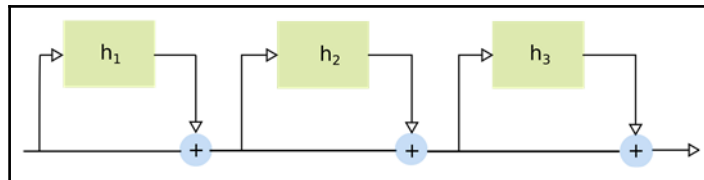
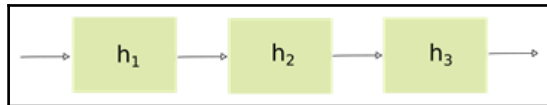
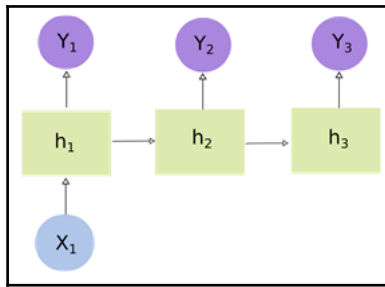
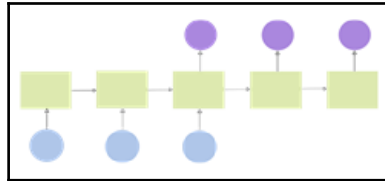
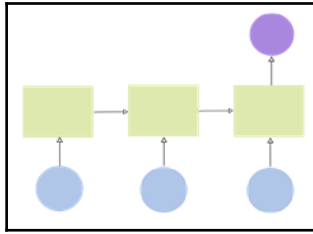
Target	Context	
fox	plan	} Training Data (Negative)
fox	son	
fox	**	
fox	**	

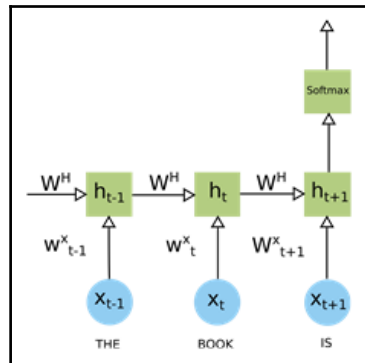
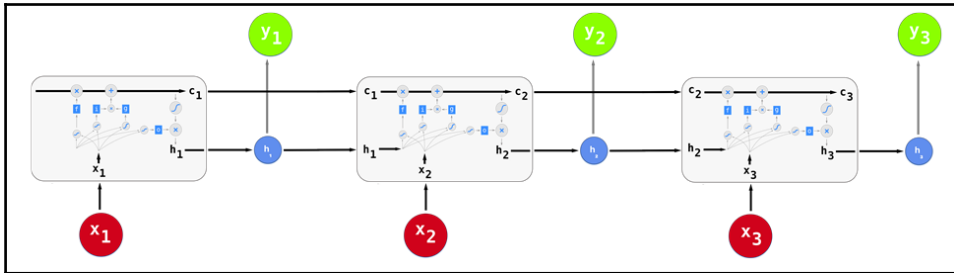
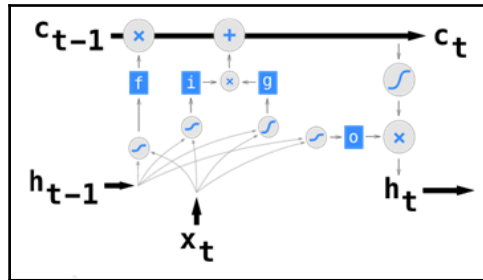
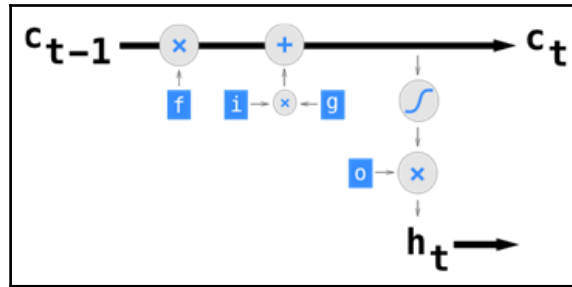


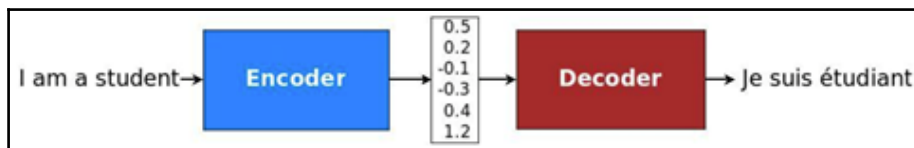
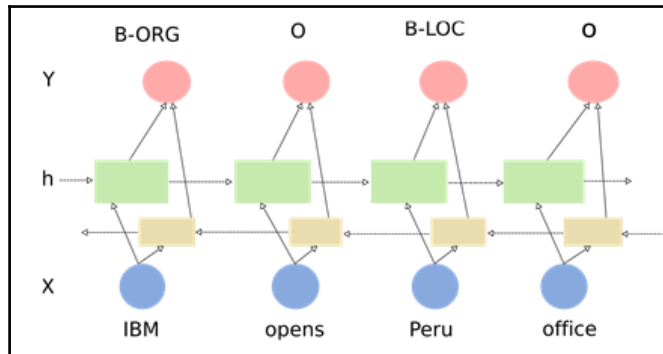
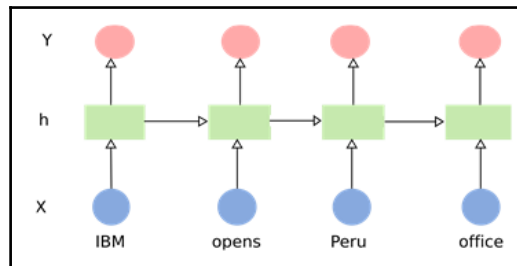
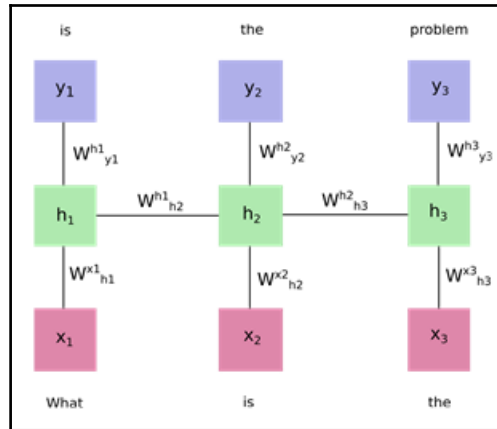


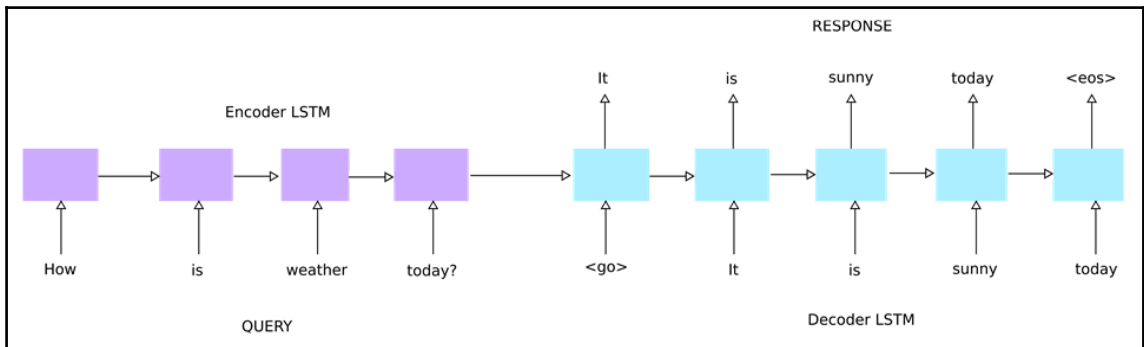
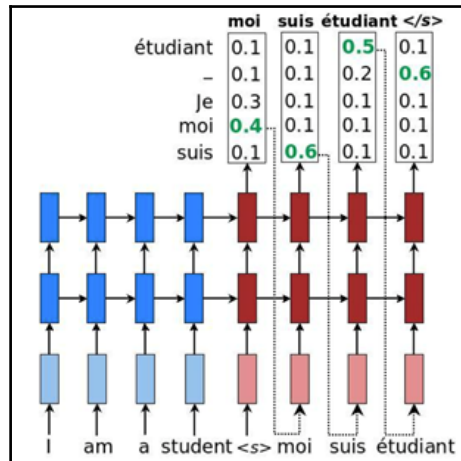
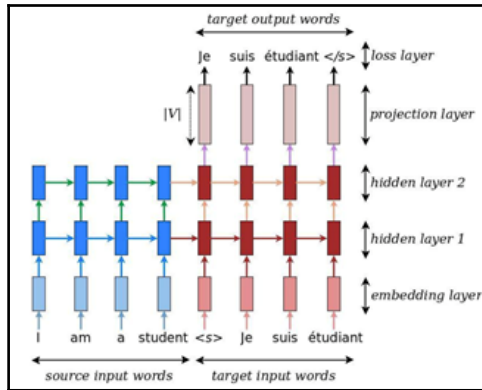
Chapter 6: Advanced Natural Language Processing



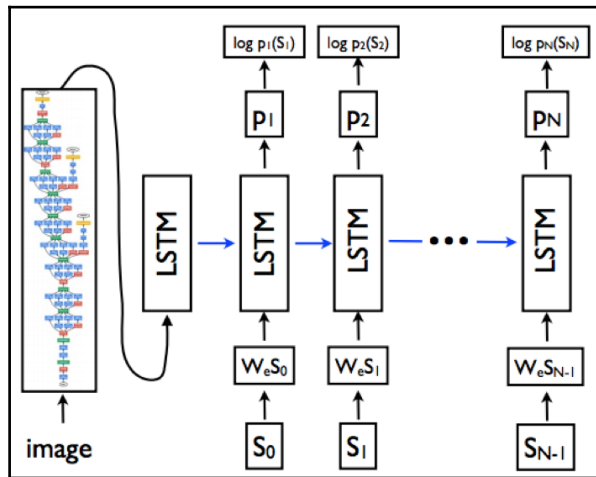
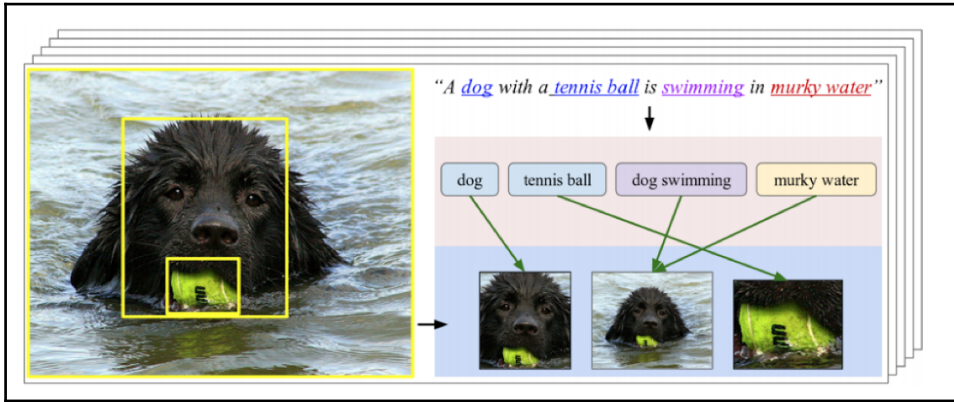




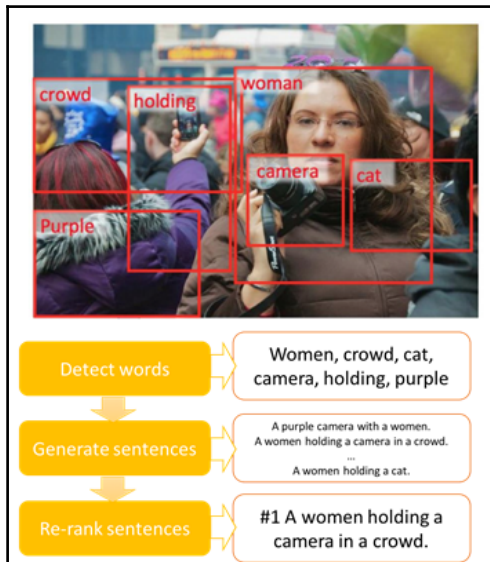




Chapter 7: Multimodality

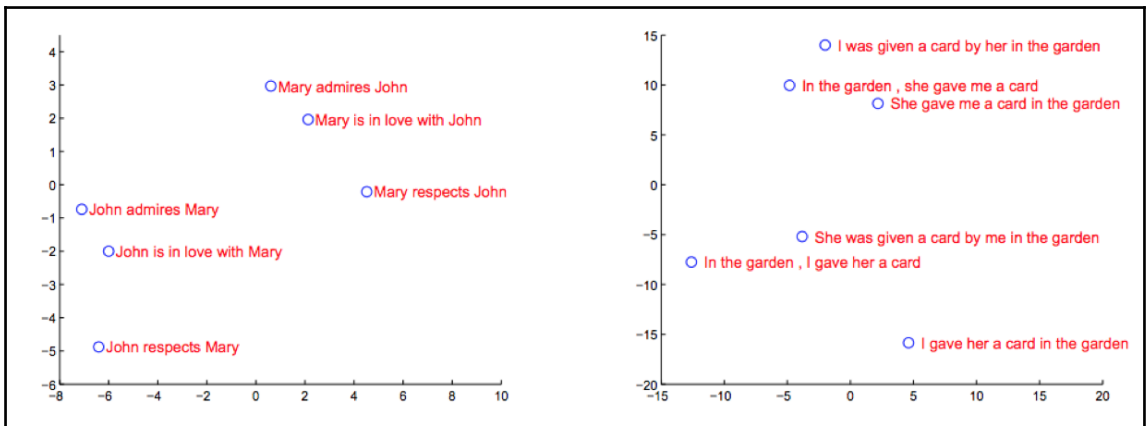
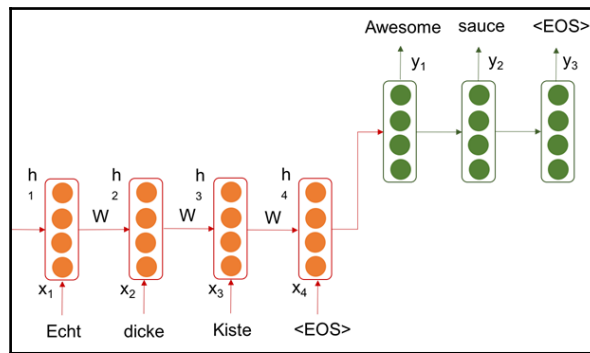


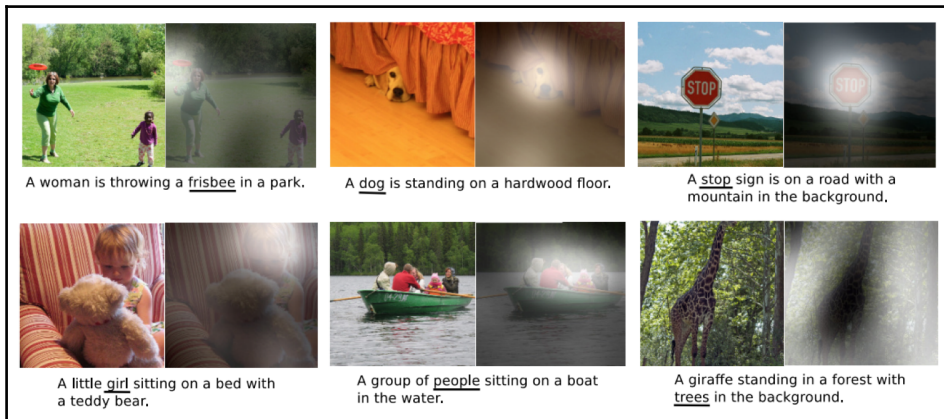
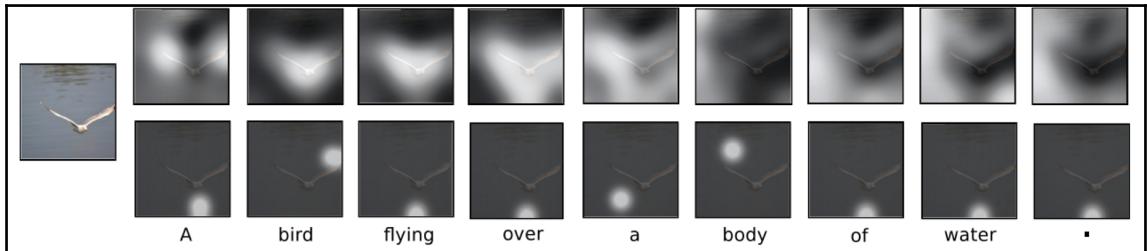
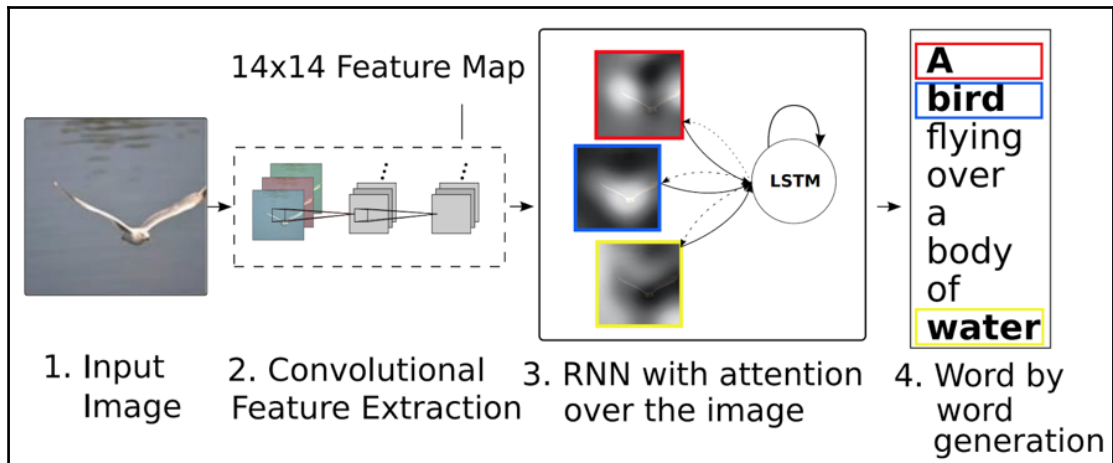
<p>A person riding a motorcycle on a dirt road.</p> 	<p>Two dogs play in the grass.</p> 	<p>A skateboarder does a trick on a ramp.</p> 	<p>A dog is jumping to catch a frisbee.</p> 
<p>A group of young people playing a game of frisbee.</p> 	<p>Two hockey players are fighting over the puck.</p> 	<p>A little girl in a pink hat is blowing bubbles.</p> 	<p>A refrigerator filled with lots of food and drinks.</p> 
<p>A herd of elephants walking across a dry grass field.</p> 	<p>A close up of a cat laying on a couch.</p> 	<p>A red motorcycle parked on the side of the road.</p> 	<p>A yellow school bus parked in a parking lot.</p> 
Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image



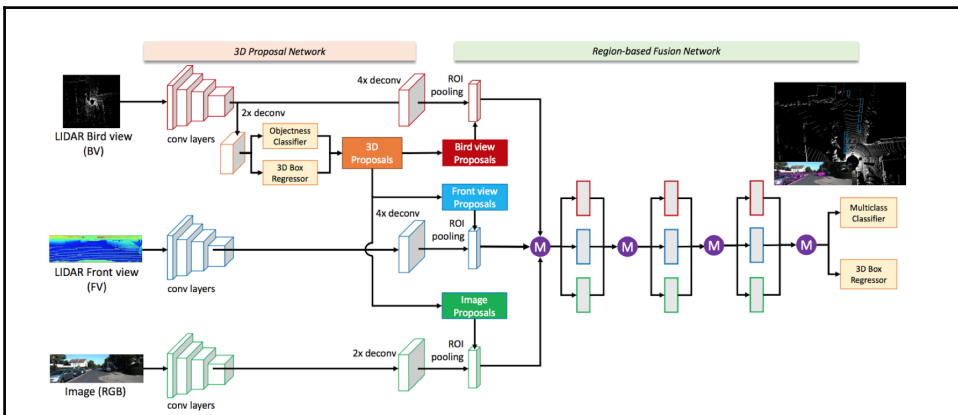
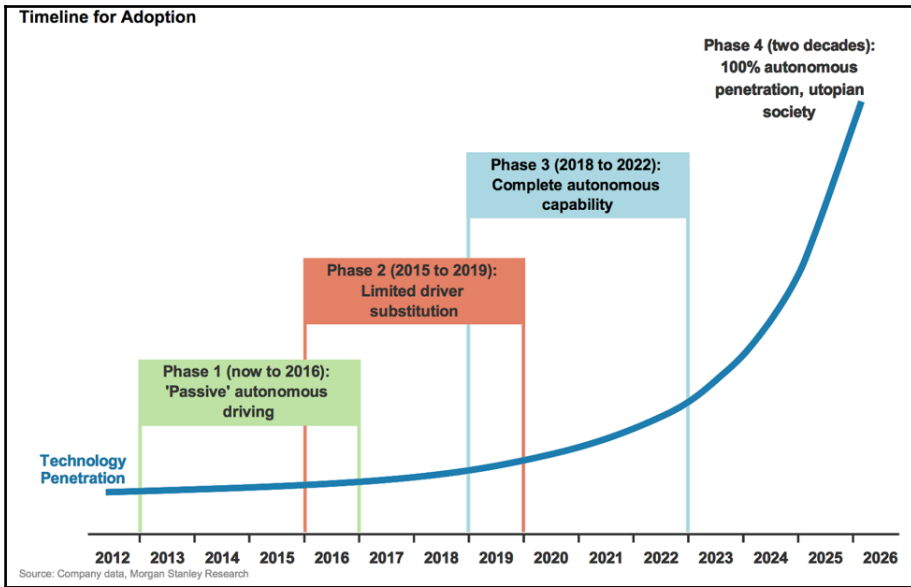
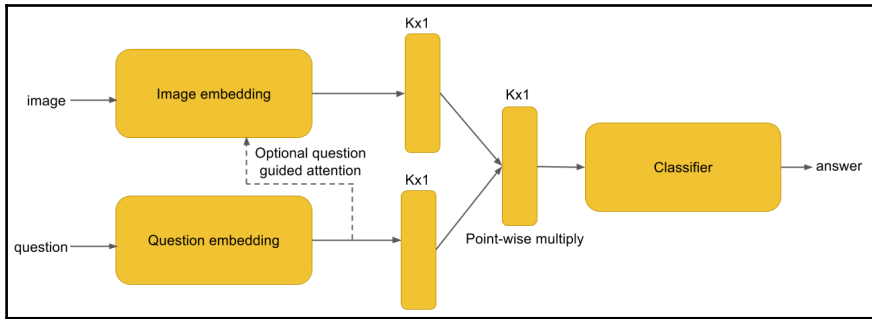


- Basketball player has fallen on the court while another grabs at the ball from out of frame.
- Two basketball players are scrambling for the ball on the court.
- Two basketball players, one on the floor, struggle to gain possession of a basketball.
- Two high school basketball players reach to grab the ball, one falling to the floor.
- Two men in uniforms playing basketball, struggle for the ball.

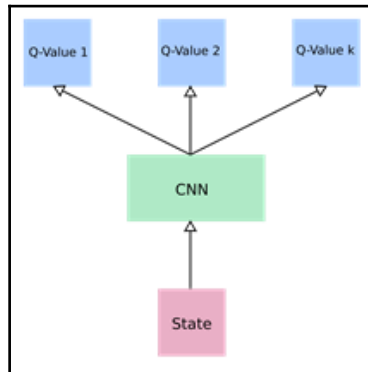
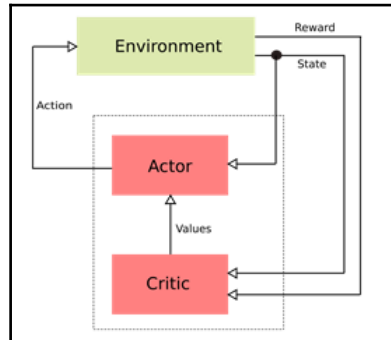
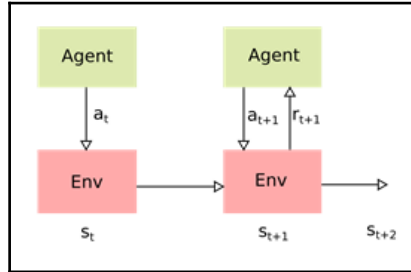


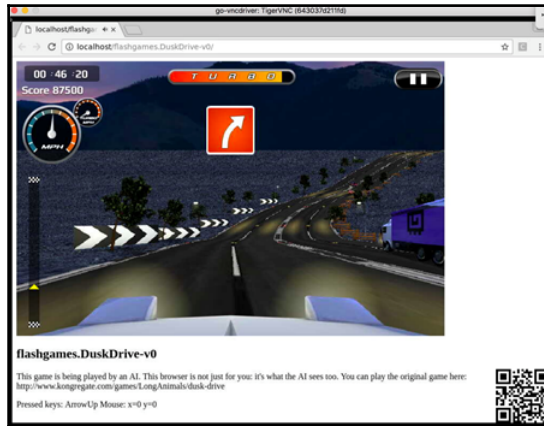
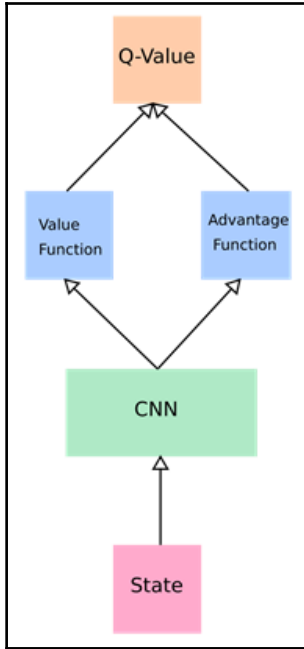






Chapter 8: Deep Reinforcement Learning



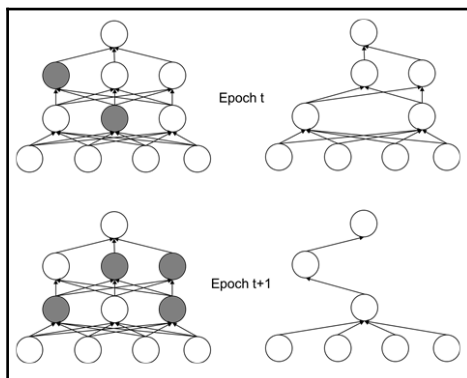


```
Next State = 4, Reward = 0.0, Q-Value = 0.549656867981
Next State = 4, Reward = 0.0, Q-Value = 0.558322191238
Next State = 4, Reward = 0.0, Q-Value = 0.561794042587
Next State = 4, Reward = 0.0, Q-Value = 0.605571866035
Next State = 10, Reward = 0.0, Q-Value = 0.527264535427
Next State = 0, Reward = 0.0, Q-Value = 0.58230805397
Next State = 9, Reward = 0.0, Q-Value = 0.662398397923
Next State = 6, Reward = 0.0, Q-Value = 0.342416584492
Next State = 12, Reward = 0.0, Q-Value = 0.00967993494123
Next State = 0, Reward = 0.0, Q-Value = 0.549962639809
Next State = 0, Reward = 0.0, Q-Value = 0.548377752304
Next State = 4, Reward = 0.0, Q-Value = 0.557400584221
Next State = 4, Reward = 0.0, Q-Value = 0.573505043983
Next State = 8, Reward = 0.0, Q-Value = 0.609502971172
Next State = 10, Reward = 0.0, Q-Value = 0.647010564804
Next State = 7, Reward = 0.0, Q-Value = 0.00918172858655
Next State = 0, Reward = 0.0, Q-Value = 0.567937016487
Next State = 4, Reward = 0.0, Q-Value = 0.596188902855
Next State = 8, Reward = 0.0, Q-Value = 0.615541934967
Next State = 0, Reward = 0.0, Q-Value = 0.555838167667
Next State = 4, Reward = 0.0, Q-Value = 0.554713010788
Next State = 6, Reward = 0.0, Q-Value = 0.30840498209
Next State = 6, Reward = 0.0, Q-Value = 0.373454988003
Next State = 0, Reward = 0.0, Q-Value = 0.520658791065
Next State = 13, Reward = 0.0, Q-Value = 0.777940273285
Next State = 15, Reward = 1.0, Q-Value = 0.00980058684945
Next State = 10, Reward = 0.0, Q-Value = 0.605933964252
Next State = 0, Reward = 0.0, Q-Value = 0.542133152485
Next State = 14, Reward = 0.0, Q-Value = 0.843689084053
Next State = 6, Reward = 0.0, Q-Value = 0.365039229393
```

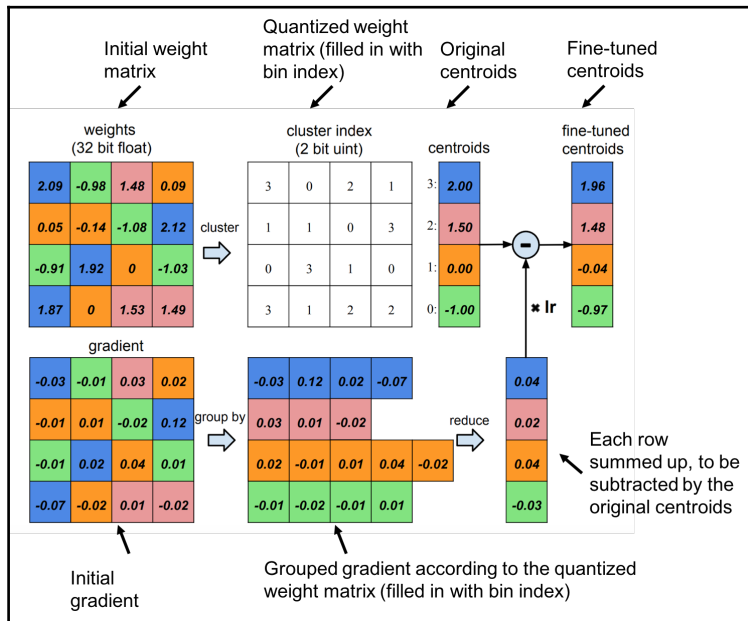
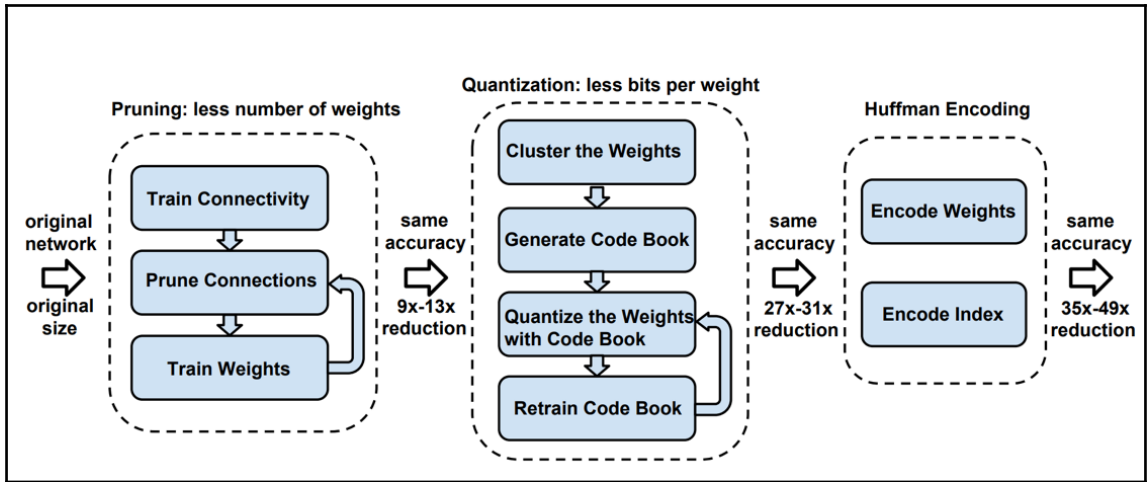
Chapter 9: Deep Learning Hacks

$$\theta^{(t+1)} = \theta^t - \epsilon_k \nabla_{\theta} L$$

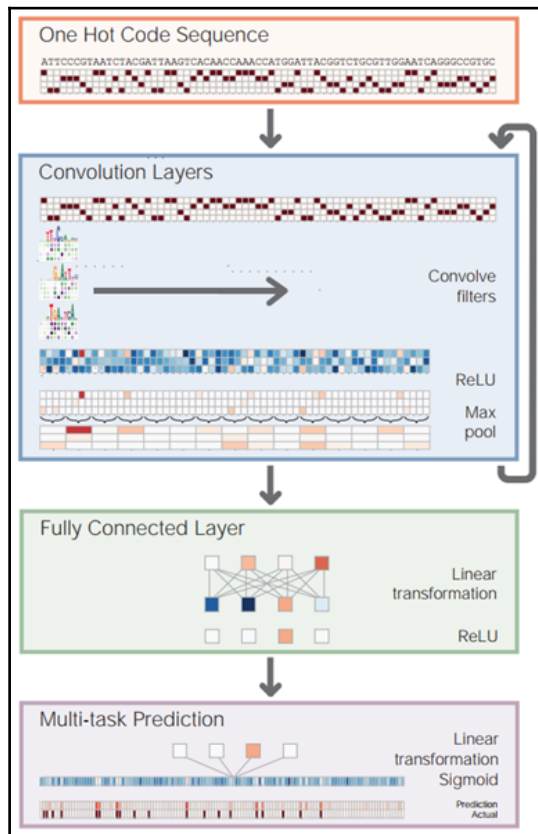
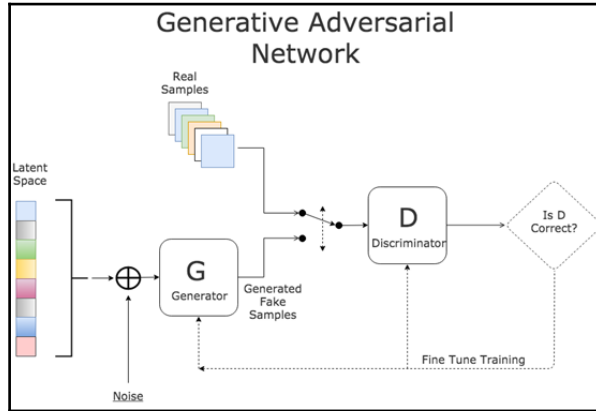
New model parameters at t+1 Old model parameters Learning rate at iteration k Gradient of loss function

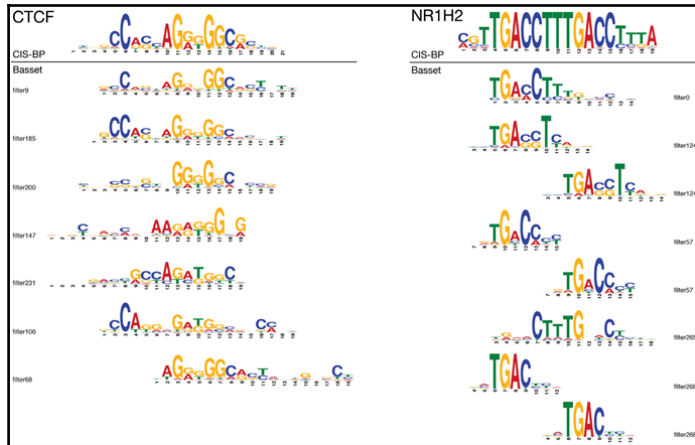


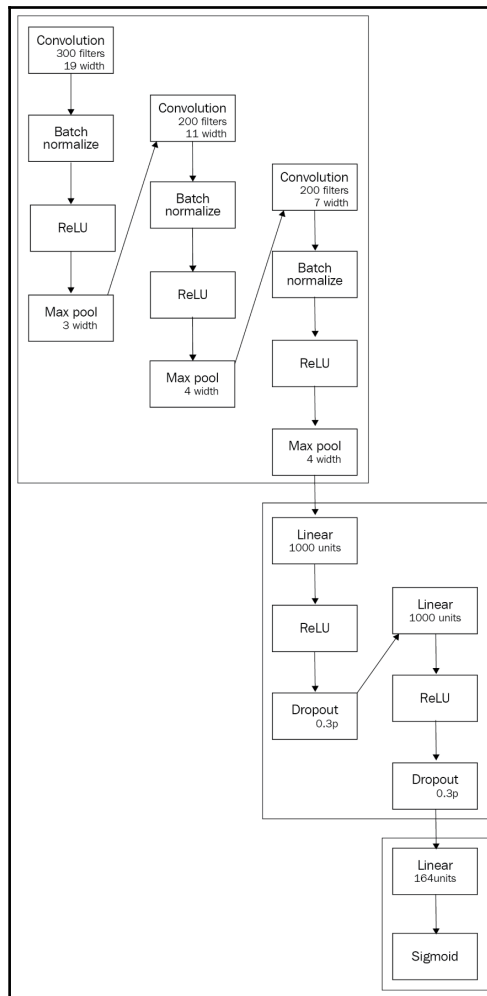
Theme Name	Description	Applications	More details
Parameter pruning and sharing	Reducing redundant parameters which are not sensitive to the performance	Convolutional layer and fully connected layer	Robust to various settings, can achieve good performance, can support both train from scratch and pre-trained model
Low-rank factorization	Using matrix/tensor decomposition to estimate the informative parameters	Convolutional layer and fully connected layer	Standardized pipeline, easily to be implemented, can support both train from scratch and pre-trained model
Transferred/compact convolutional filters	Designing special structural convolutional filters to save parameters	Only for convolutional layer	Algorithms are dependent on applications, usually achieve good performance only support train from scratch
Knowledge distillation	Training a compact neural network with distilled knowledge of a large model	Convolutional layer and fully connected layer	Model performances are sensitive to applications and network structure only support train from scratch

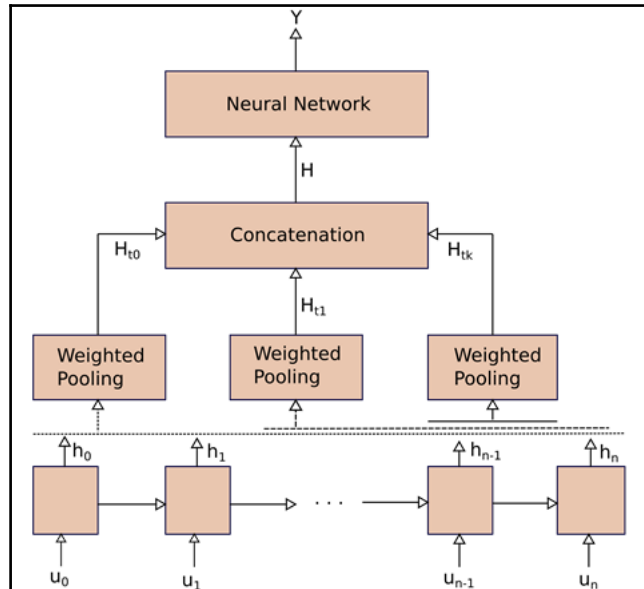
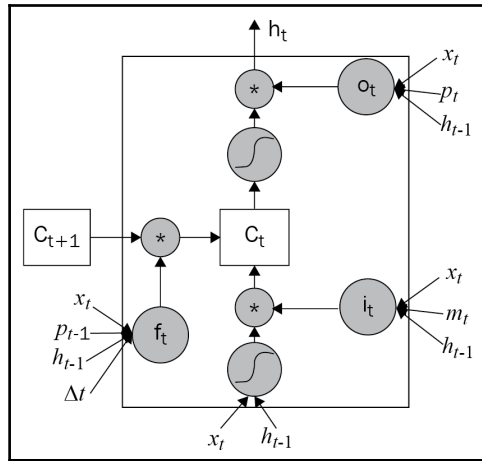


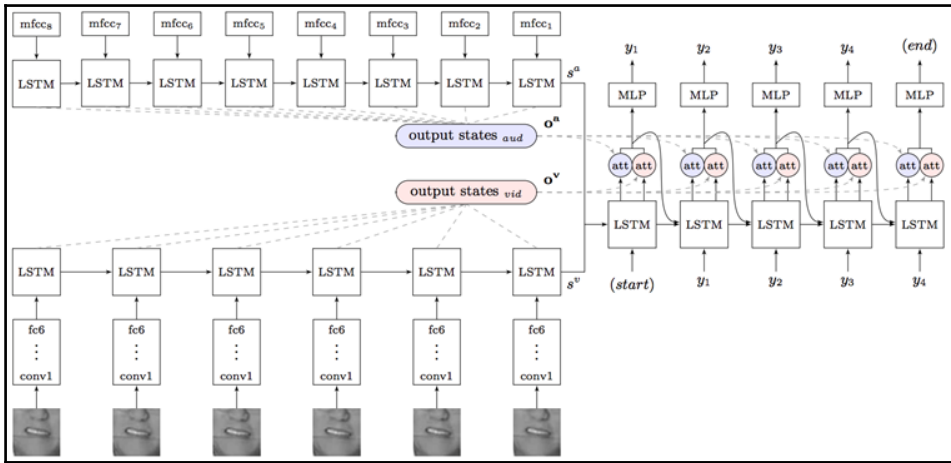
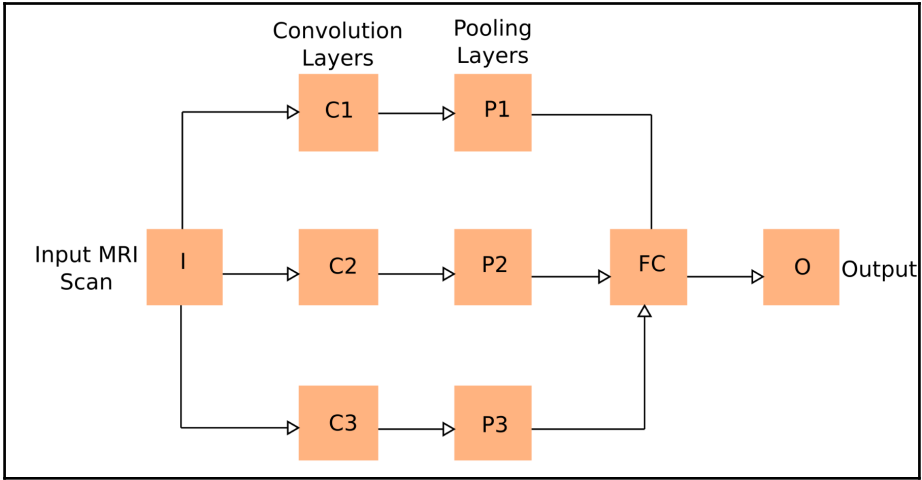
Chapter 10: Deep Learning Trends

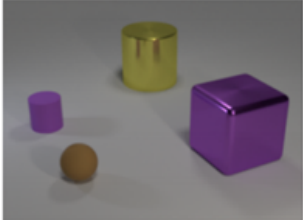
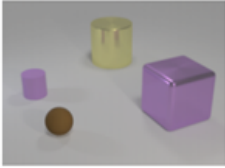
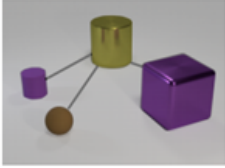


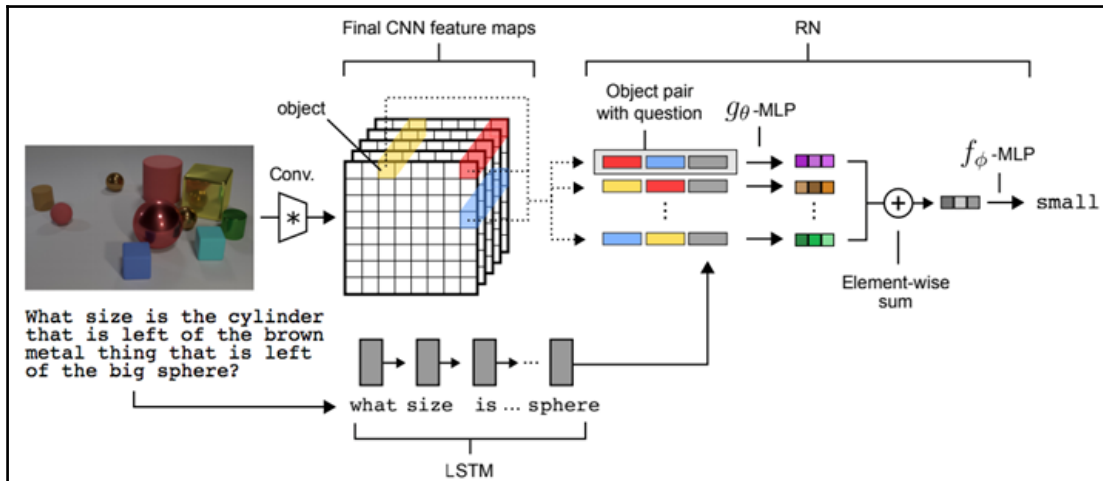


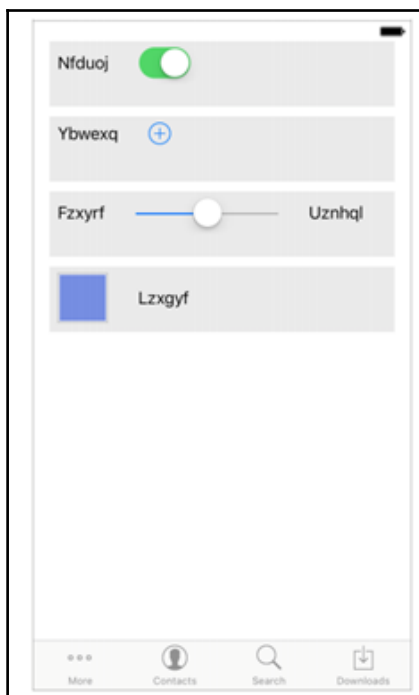






<p>Original Image:</p> 	<p>Non-relational question: What is the size of the brown sphere?</p>	
	<p>Relational question: Are there any rubber things that have the same size as the yellow metallic cylinder?</p>	





```
stack {
  row {
    label, switch
  }
  row {
    label, btn-add
  }
  row {
    label, slider, label
  }
  row {
    img, label
  }
}
footer {
  btn-more, btn-contact, btn-search, btn-download
}
```

