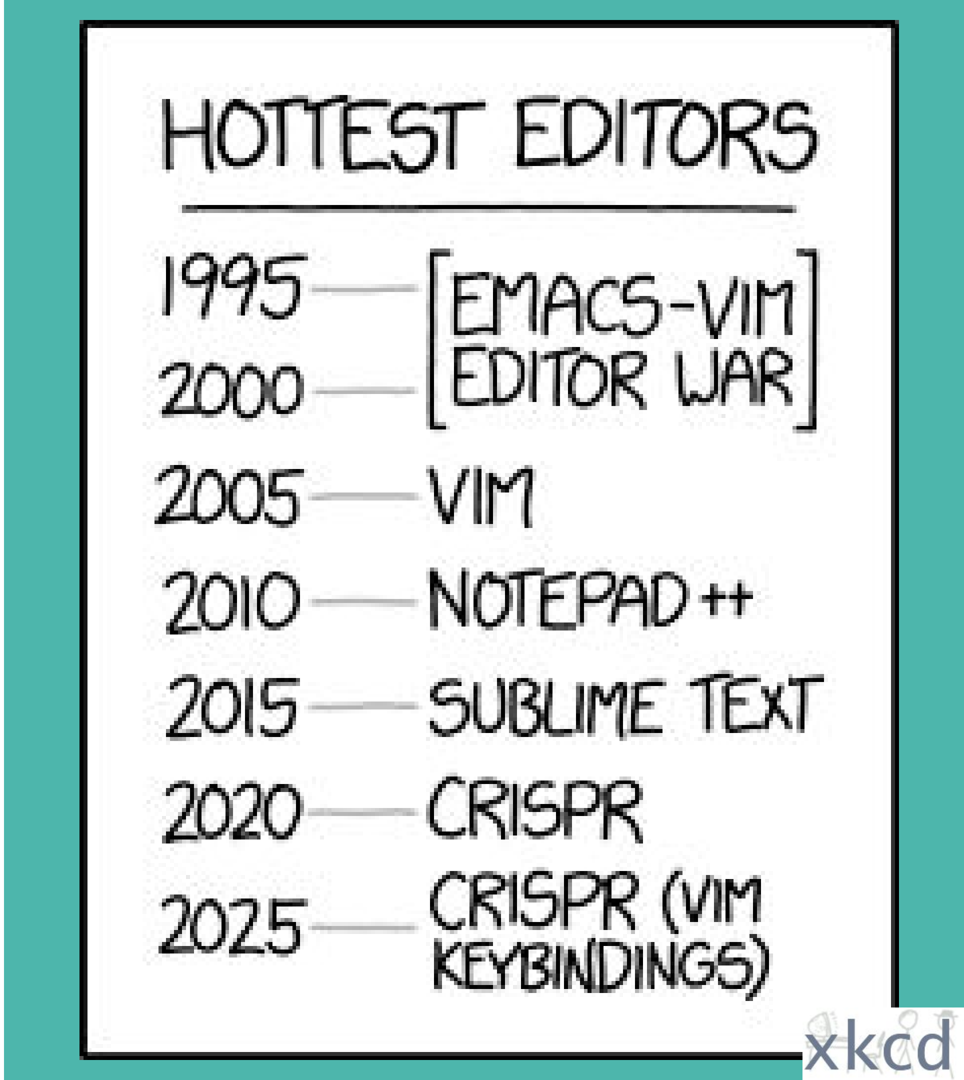
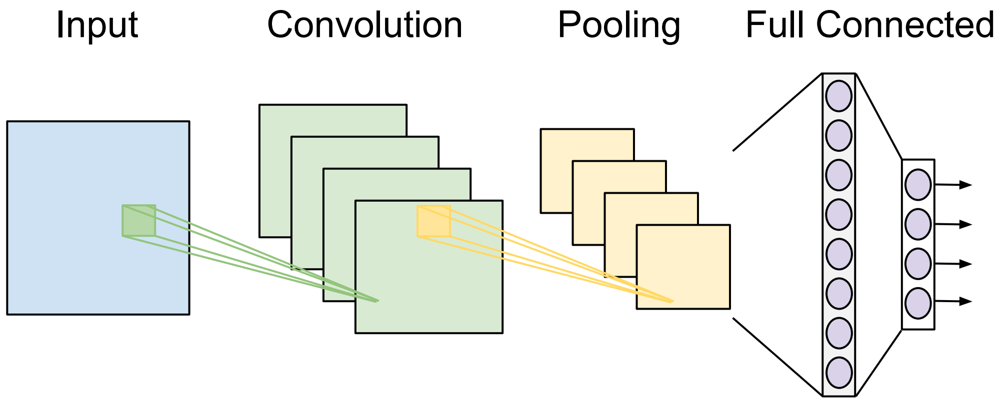
Recent research concepts in deep learning

Presented by:
Ryan Hausen,
Molly Zhang, and
David Parks



Convolutional Neural Networks

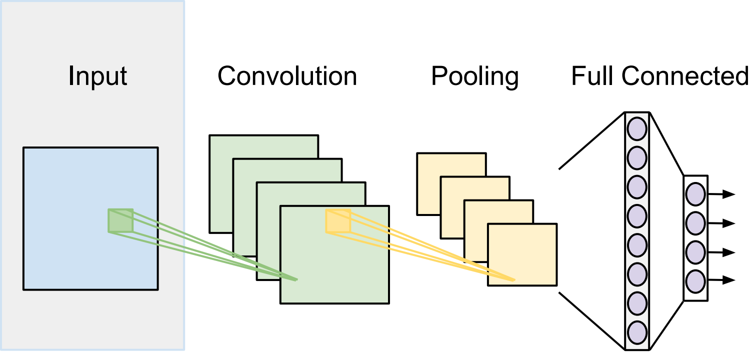
- A type of neural network that takes advantage of the spatial information stored in data
- Comprised of convolution, pooling, and fully connected layers
- Very popular in image datasets

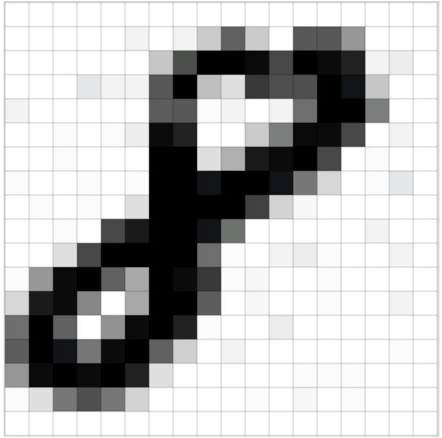


Convolutional Neural Networks

Input data should have spatial value.

Doesn't necessarily need to be an image.





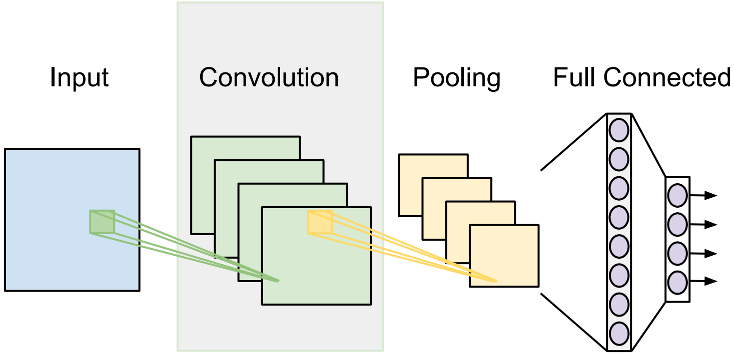
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	12	0	11	39	137	37	0	152	147	84	0	0	0
0	0	1	0	0	0	41	160	250	255	235	162	255	238	206	11	13	0
0	0	0	16	9	9	150	251	45	21	184	159	154	255	233	40	0	0
10	0	0	0	0	0	145	146	3	10	0	11	124	253	255	107	0	0
0	0	3	0	4	15	236	216	0	0	38	109	247	240	169	0	11	0
1	0	2	0	0	0	253	253	23	62	224	241	255	164	0	5	0	0
6	0	0	4	0	3	252	250	228	255	255	234	112	28	0	2	17	0
0	2	1	4	0	21	255	253	251	255	172	31	8	0	1	0	0	0
0	0	4	0	163	225	251	255	229	120	0	0	0	0	0	11	0	0
0	0	21	162	255	255	254	255	126	6	0	10	14	6	0	0	9	0
3	79	242	255	141	66	255	245	189	7	8	0	0	5	0	0	0	0
26	221	237	98	0	67	251	255	144	0	8	0	0	7	0	0	11	0
125	255	141	0	87	244	255	208	3	0	0	13	0	1	0	1	0	0
145	248	228	116	235	255	141	34	0	11	0	1	0	0	0	1	3	0
85	237	253	246	255	210	21	1	0	1	0	0	6	2	4	0	0	0
6	23	112	157	114	32	0	0	0	0	2	0	8	0	7	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

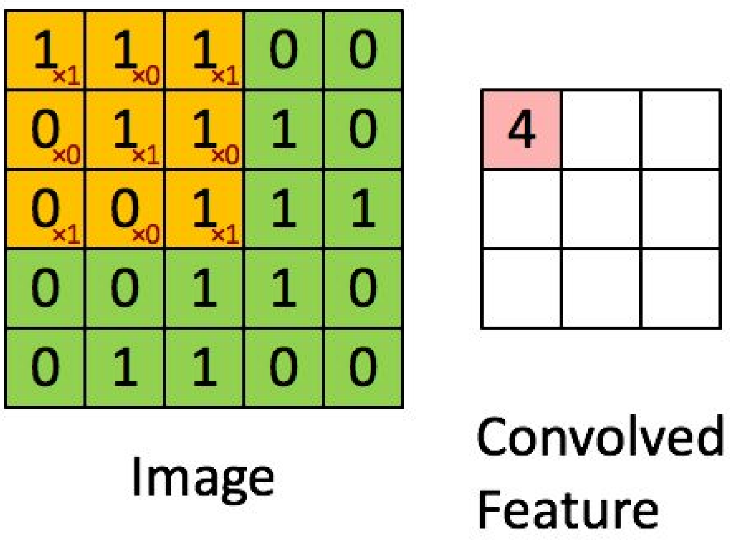
https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

Convolutional Neural Networks

The Convolution layer uses kernels to extract patterns from the image.

The kernels are learned during training



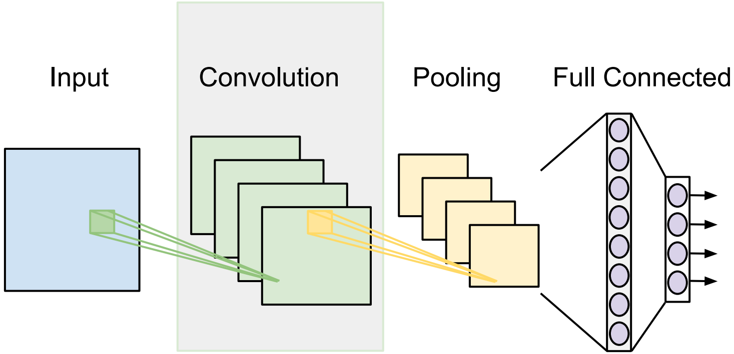


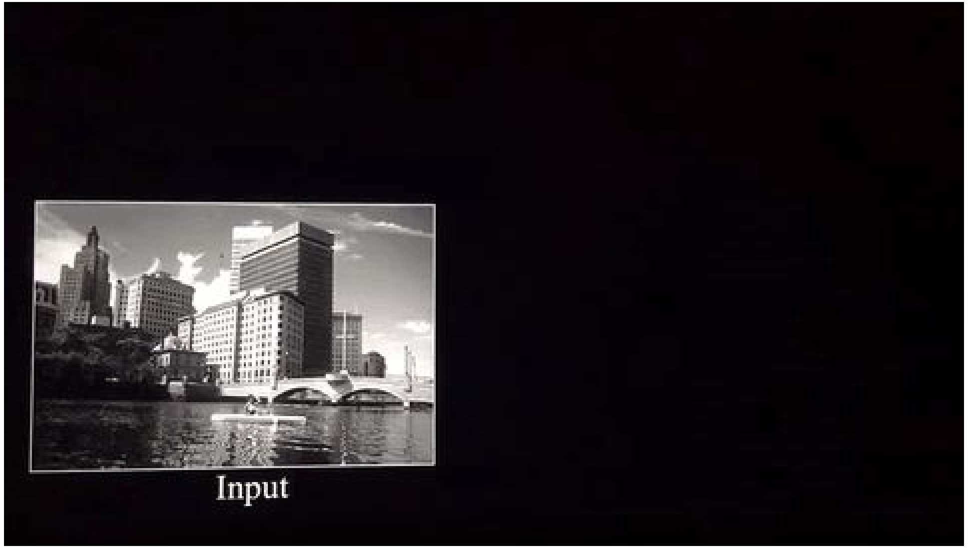
https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

Convolutional Neural Networks

The Convolution layer uses kernels to extract patterns from the image.

The kernels are learned during training

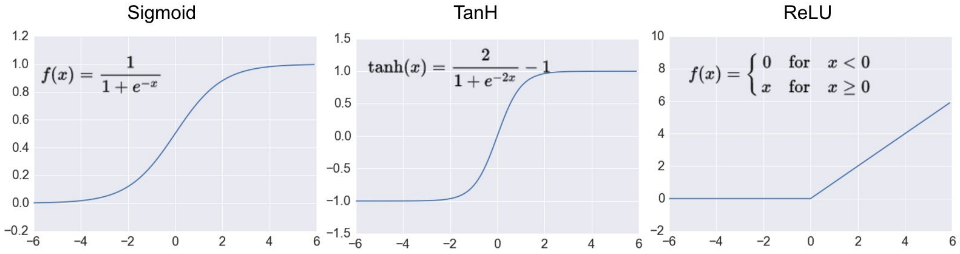


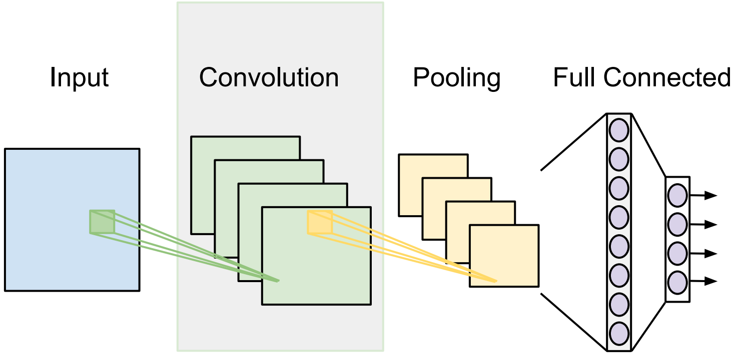


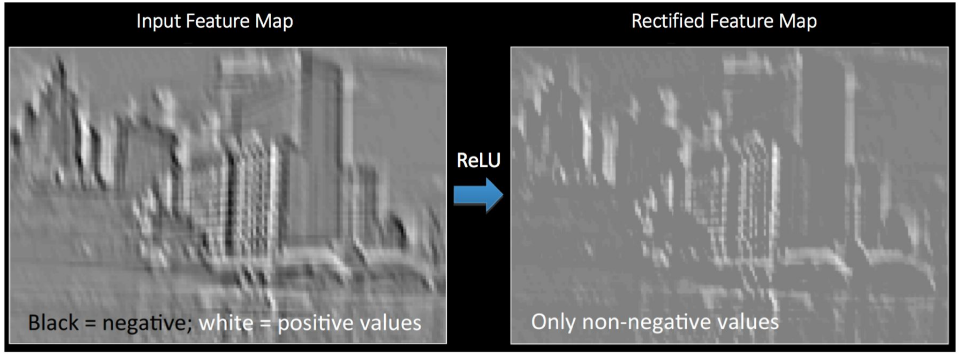
https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

Convolutional Neural Networks

After the convolution of the kernel the values are passed through a nonlinearity.



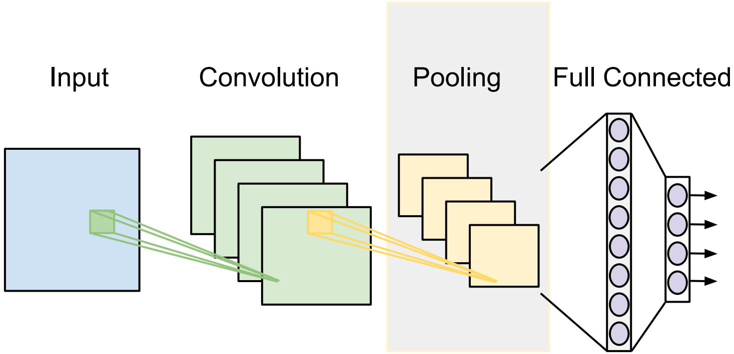


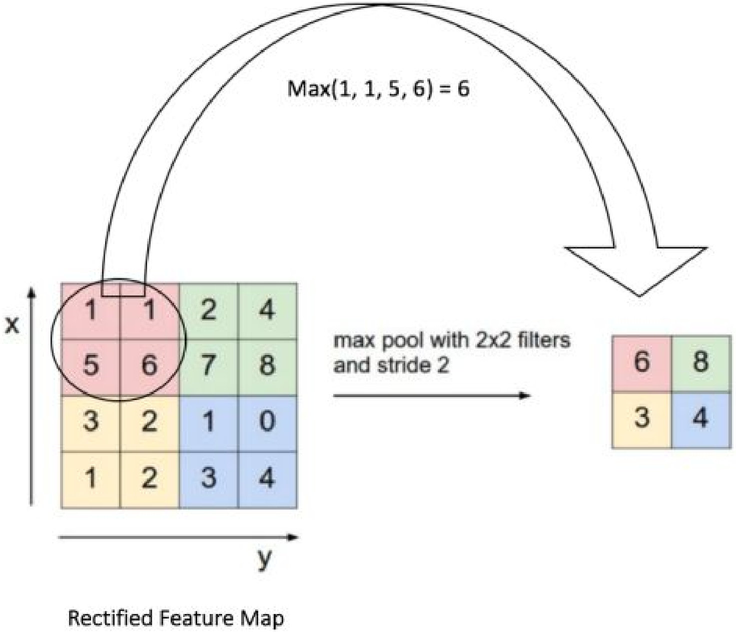


https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

Convolutional Neural Networks

Optionally, outputs from the convolution layer can be further distilled by a pooling layer.

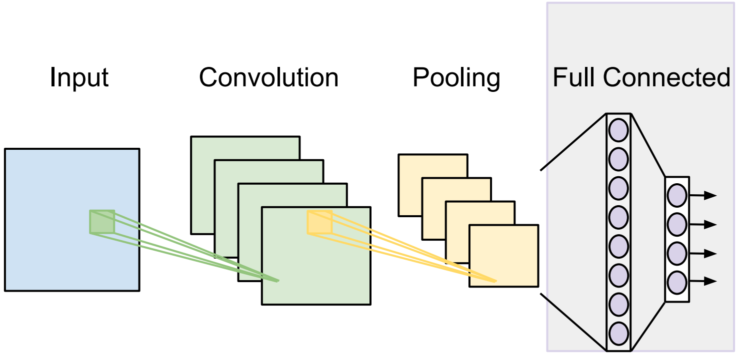


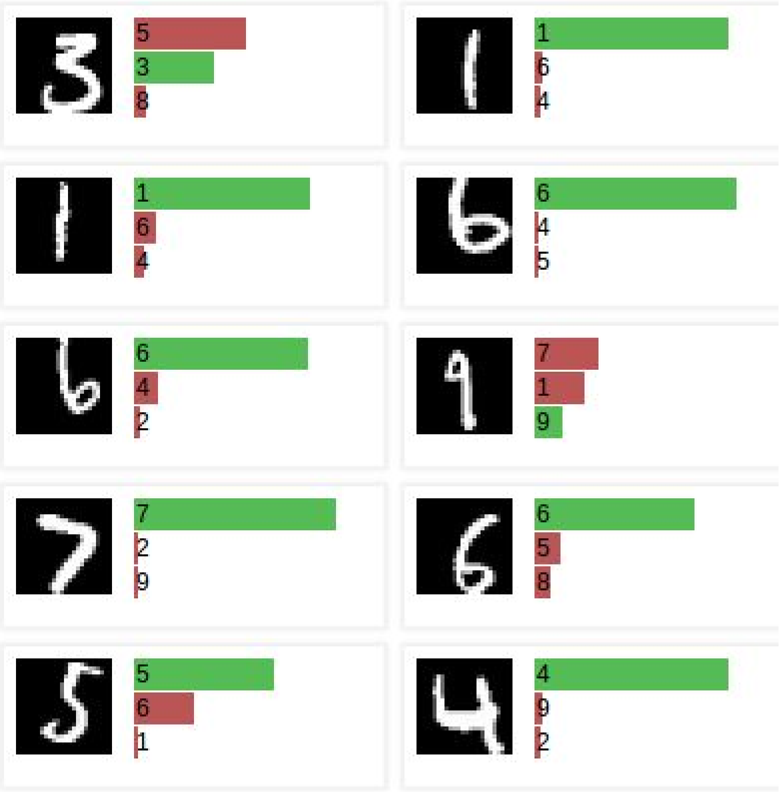


https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

Convolutional Neural Networks

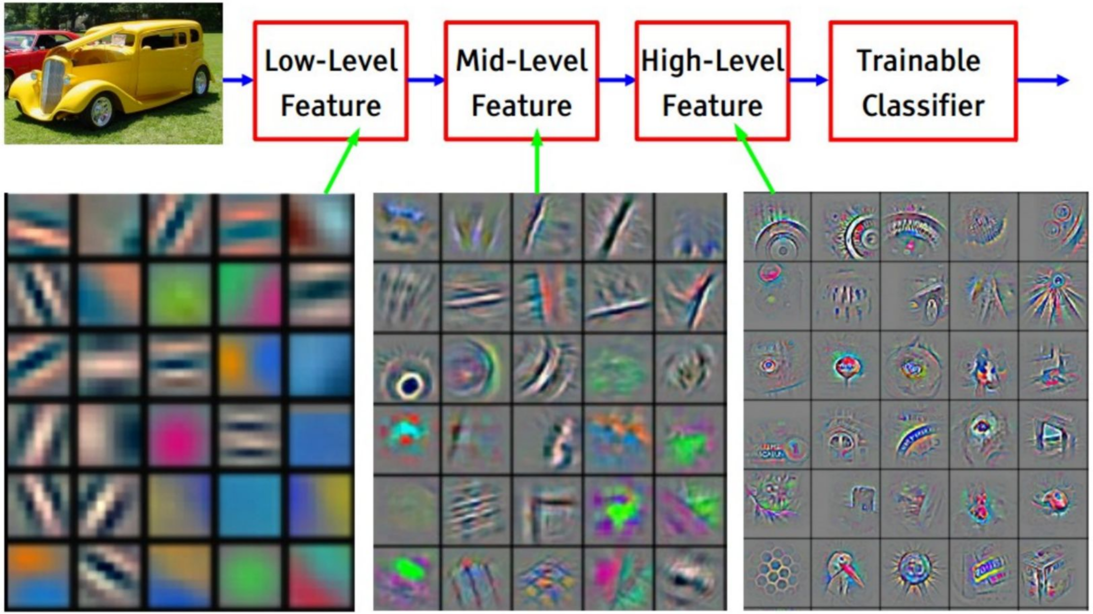
Finally, the processed input data is processed by a/some fully connected layer/s and a classification is output.





Example

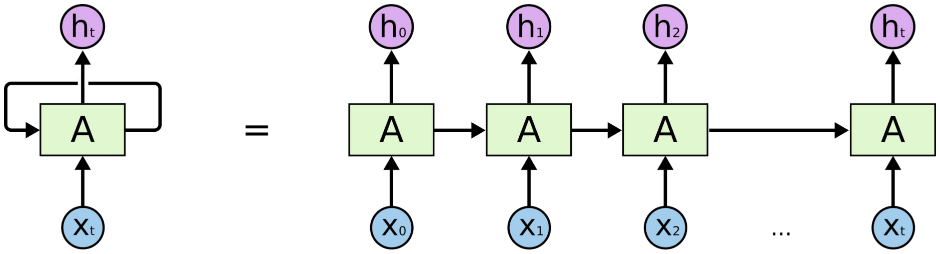
[From recent Yann LeCun slides]



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

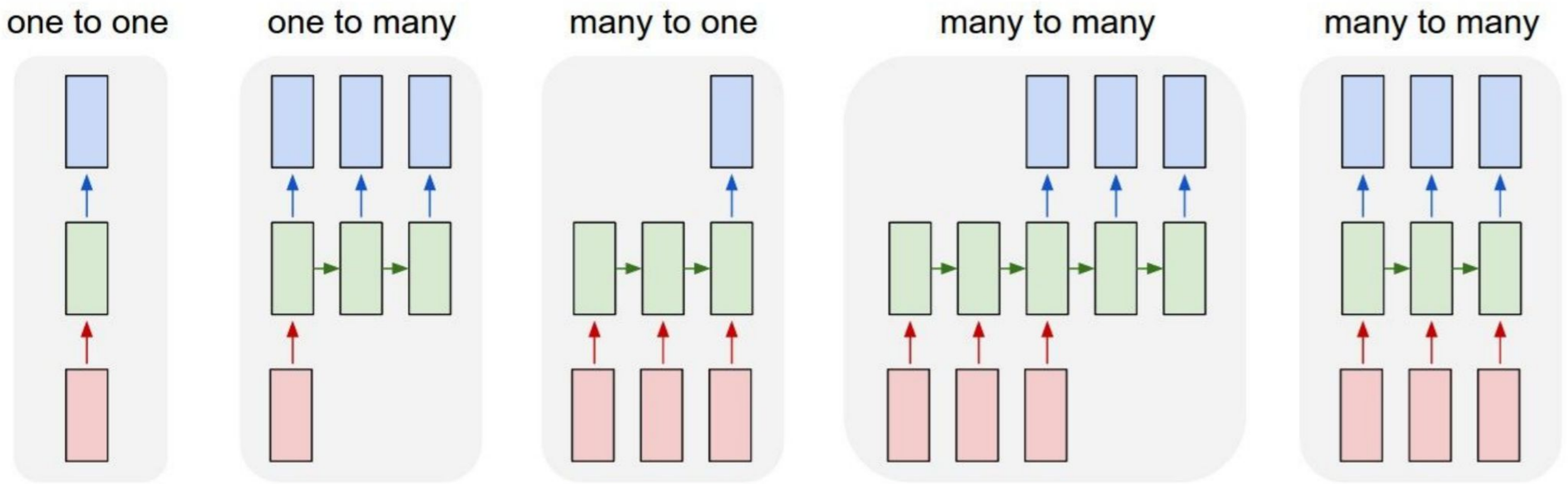
Recurrent Neural Networks

- A type of neural network that takes advantage of the sequential nature of data
- Can be very compact
- Utilizes shared weights over time steps



http://colah.github.io/posts/2015-08-Understanding-LSTMs

Recurrent Neural Networks



Vanilla	Image Captioning Sequence of words	Sentiment Classification Words -> Sentiment	Translation seq of words -> seq of words	Video classification frame level

* http://cs231n.stanford.edu/slides/winter1516_lecture10.pdf

Example

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

Character level RNN

Trained on Shakespeare

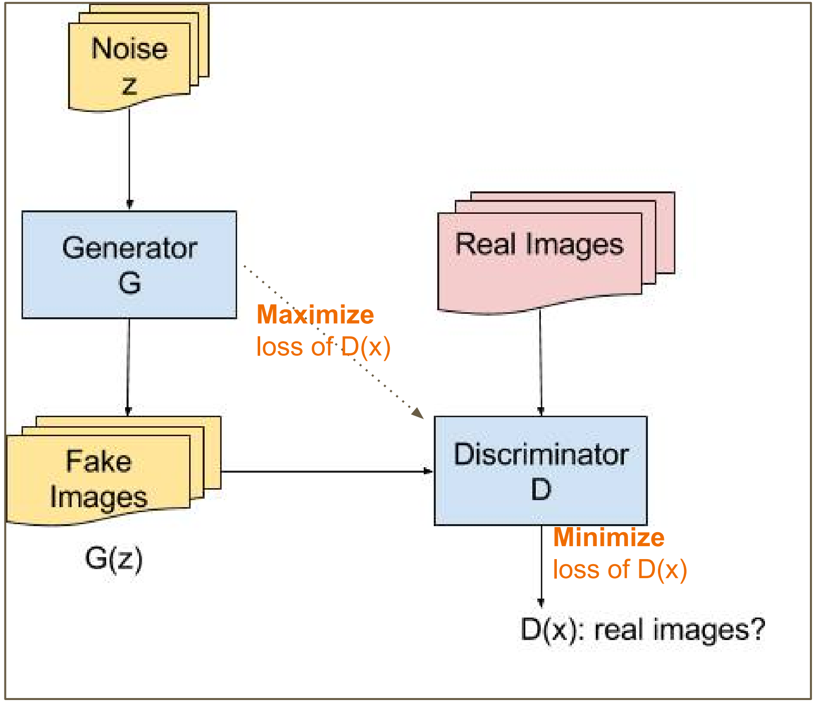
```
VIOLA:
Why, Salisbury must find his flesh and thought
That which I am not aps, not a man and in fire,
To show the reining of the raven and the wars
To grace my hand reproach within, and not a fair are hand,
That Caesar and my goodly father's world;
When I was heaven of presence and our fleets,
We spare with hours, but cut thy council I am great,
Murdered and by thy master's ready there
My power to give thee but so much as hell:
Some service in the noble bondman here,
Would show him to her wine.

KING LEAR:
O, if you were a feeble sight, the courtesy of your law,
Your sight and several breath, will wear the gods
With his heads, and my hands are wonder'd at the deeds,
So drop upon your lordship's head, and your opinion
Shall be against your honour.
```

https://karpathy.github.io/2015/05/21/rnn-effectiveness/

This page left blank intentionally

Generative Adversarial Network (GAN)



A Framework of generative model via an adversarial process

- Generator: generative "fake" image
- Discriminator Goal: Distinguish real images from fake image. (**minimizing** discriminative loss)
- Generator Goal: create images that make the discriminator think they are real (**maximizing** discriminator mistake)
- A two-player minmax game

Generative Adversarial Network

Objective Function

$$\min_G \max_D V(D, G)$$

$$= \min_G \max_D \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

- D: Discrimination model
- G: Generative model

$D(x)$: the probability that x came from real data rather than generated data.

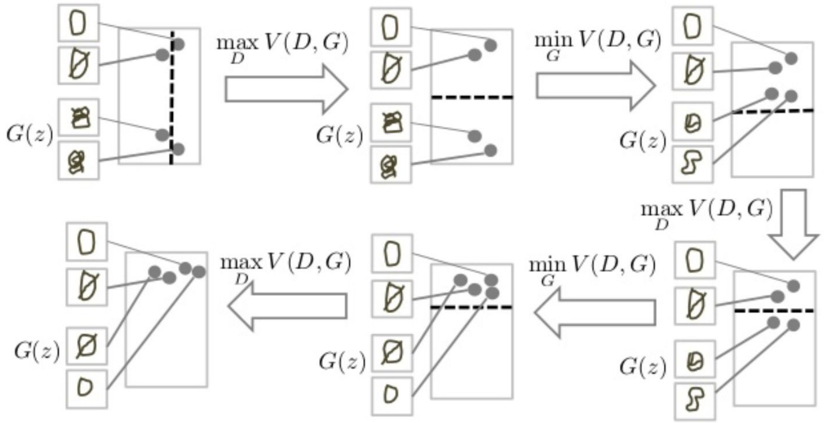
$$D : \Psi \rightarrow [0, 1]$$

$G(z)$: is a differential function to generate data from $z \sim p_z$.

$$G : \Phi \rightarrow \Psi$$

Ψ and Φ refers to the data space and generating space.

Training GAN

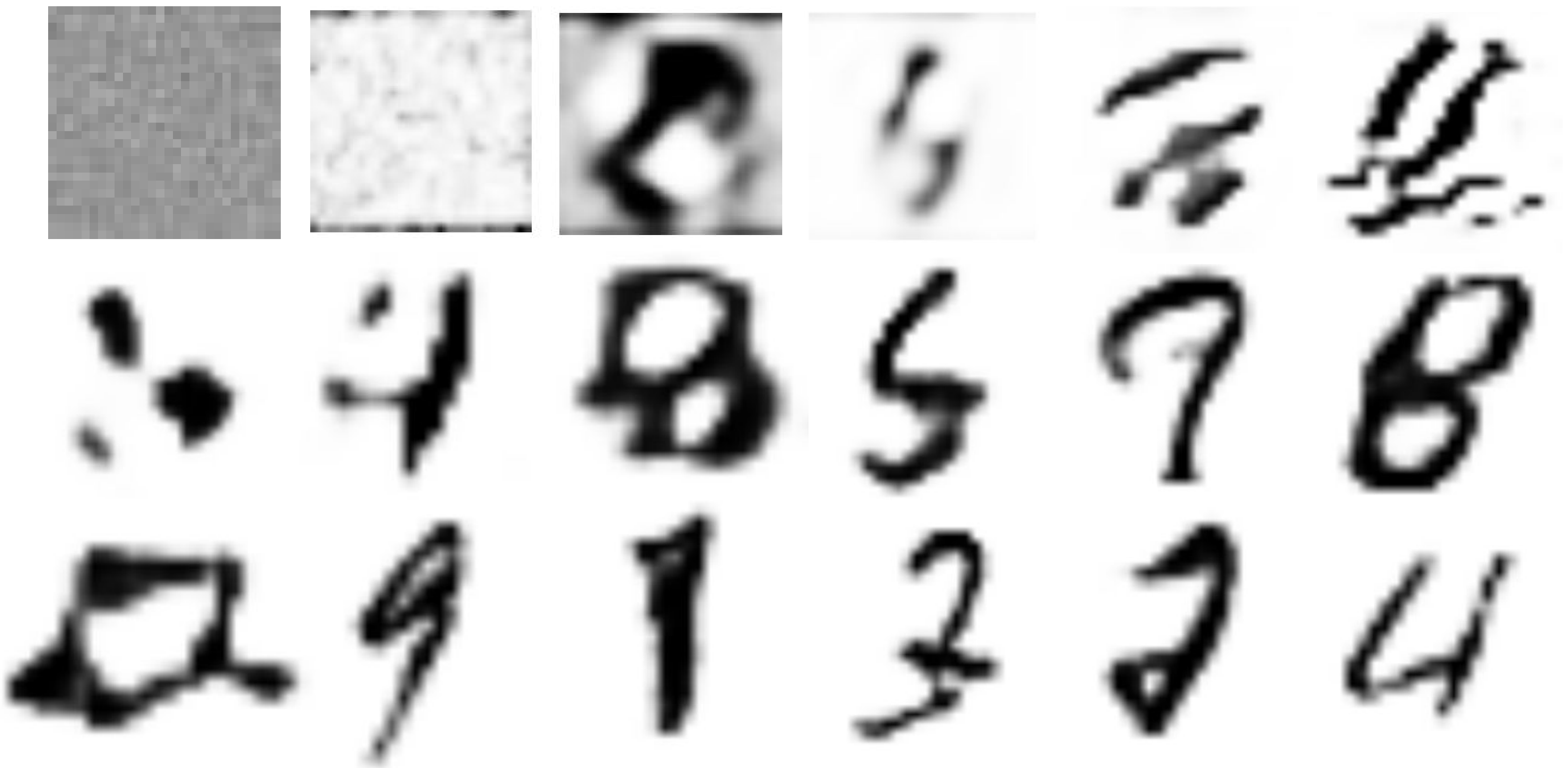


MNIST example: generating handwritten digits

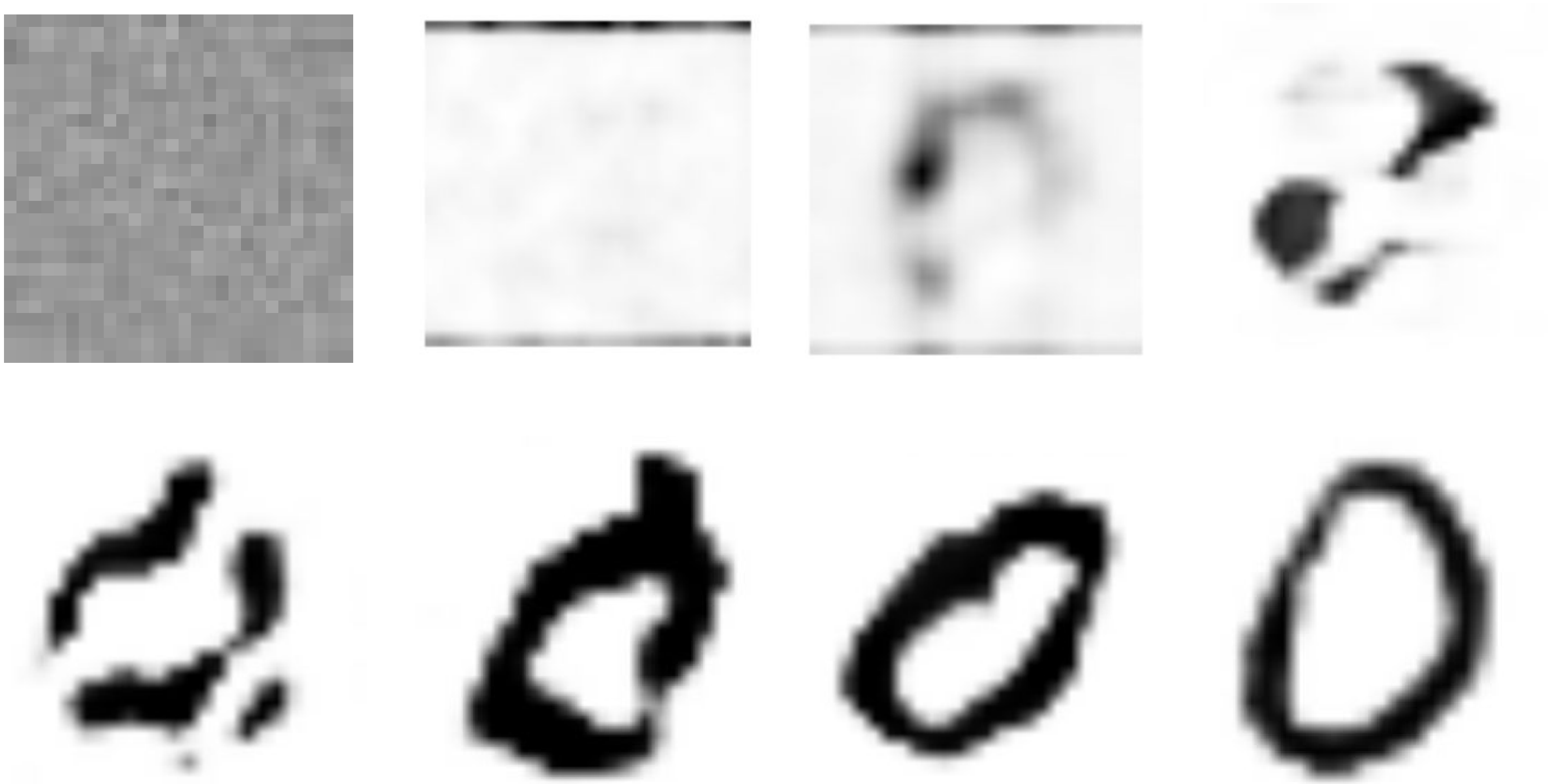
Digits written by human

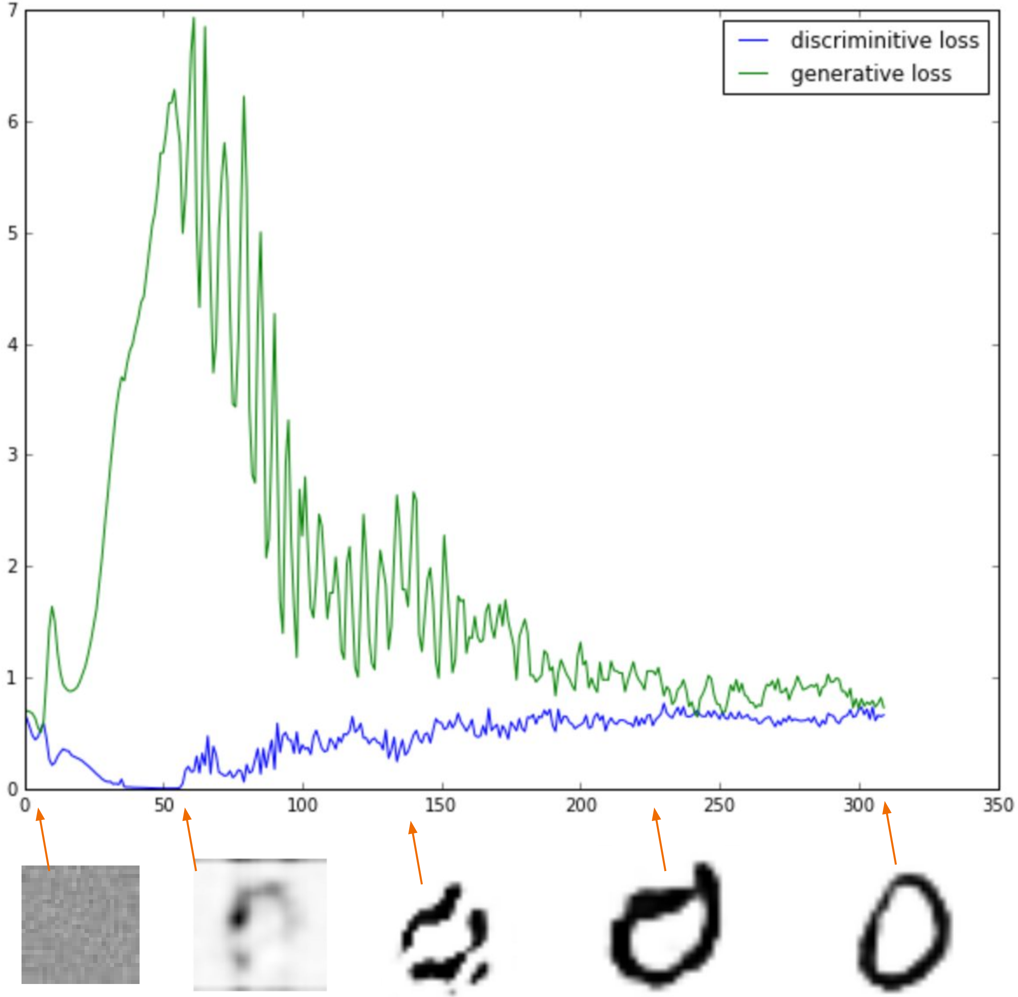
Digits generated with GAN

generating all 10 digits over 5000 iterations



generating just zeros over 300 iterations





Generative Loss and Discriminative Loss Over time

In the beginning:

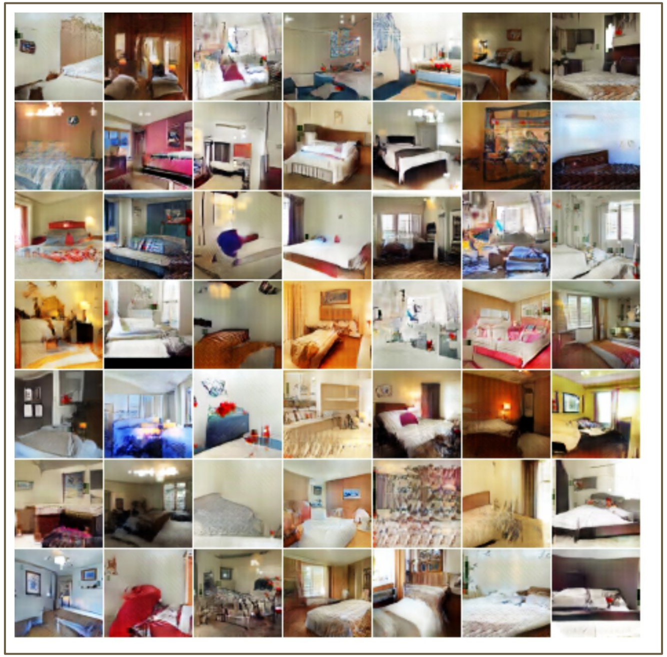
- it's easy to tell real and fake images apart (low discriminative loss)
- Easy for generator to fool discriminator (high generator loss)

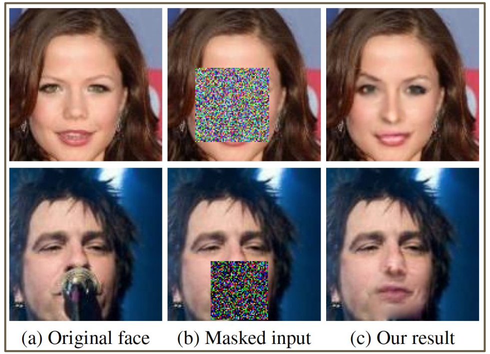
Towards the end:

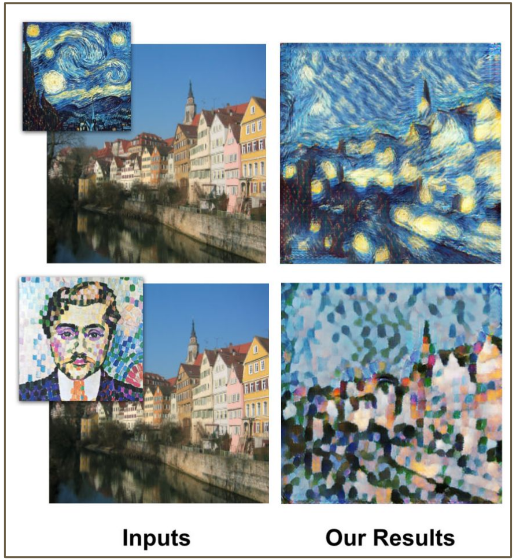
- discriminator ~0.5 loss (can't tell real/fake apart)
- generator ~0.5 loss (hard to fool discriminator)

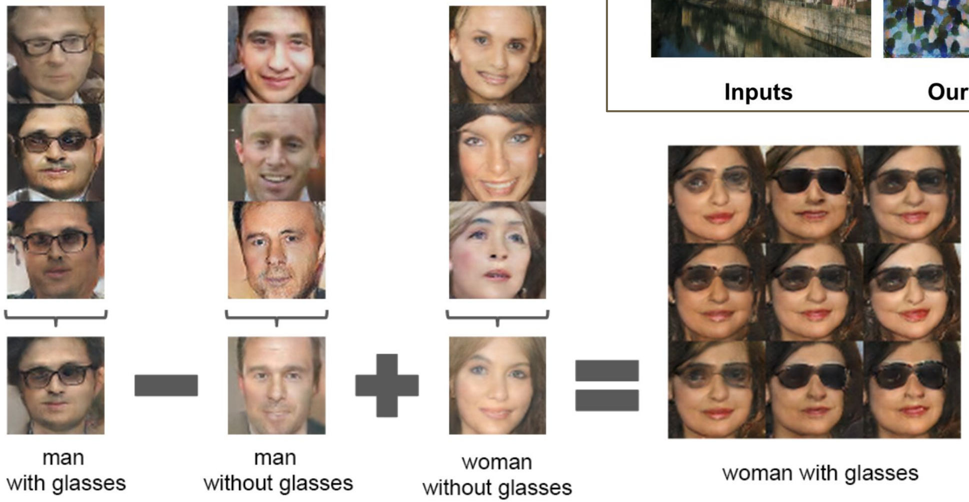
https://github.com/MollyZhang/MNIST-GAN

Applications of GAN 1: Generate, complete images, texture transformation









Papers:
https://arxiv.org/abs/1609.03126
https://arxiv.org/abs/1704.05838
https://arxiv.org/abs/1604.04382
https://arxiv.org/abs/1511.06434

Applications of GAN 2 - Generate music

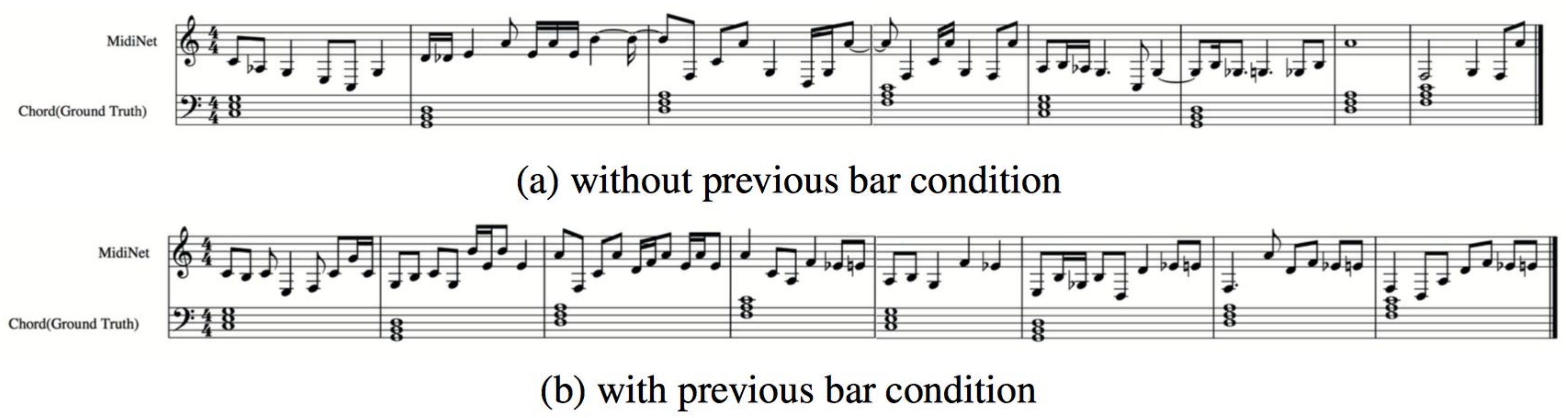


Figure 2: A sample of 8 bars generated by MidiNet.

Data: 1,022 Midi tabs with chord and melody information in separate Midi channels

The music: https://richardyang40148.github.io/TheBlog/midinet_arxiv_demo.html

Paper: https://arxiv.org/abs/1703.10847

Applications of GAN 3 - Generating biological data?

S7513 - APPLICATIONS OF GENERATIVE ADVERSARIAL NETWORKS TO DRUG DISCOVERY IN ONCOLOGY AND INFECTIOUS DISEASES

Polina Mamoshina - Sr. Research Scientist, Pharmaceutical Artificial Intelligence, Insilico Medicine, Inc

Artur Kadurin - Chief AI Officer, Insilico Medicine, Inc

Alex Zhavoronkov - CEO, Insilico Medicine, Inc

Recent advances in deep learning and specifically in generative adversarial networks have demonstrated surprising results in generating new images and videos upon request, even using natural language as input. We'll present the first application of generative adversarial autoencoders (AAE) for generating novel molecules with a defined set of parameters. In the first proof of concept experiment, we developed a seven-layer AAE architecture with the latent middle layer serving as a discriminator. As an input and output, the AAE uses a vector of binary fingerprints and concentration of the molecule. In the latent layer, we also introduced a neuron responsible for growth inhibition percentage, which, when negative, indicates the reduction in the number of tumor cells after the treatment. To train the AAE, we used the NCI-60 cell line assay data for 6252 compounds profiled on MCF-7 cell line. The output of the AAE was used to screen 72 million compounds in PubChem and select candidate molecules with potential anti-cancer properties. This approach is a proof of concept of an artificially intelligent drug discovery engine, where AAEs are used to generate new molecular fingerprints with the desired molecular properties. We'll also present the applications of this approach to discovering new anti-infective drugs and present the roadmap for generating drugs for rare diseases and even for individual patients.

Session Schedule

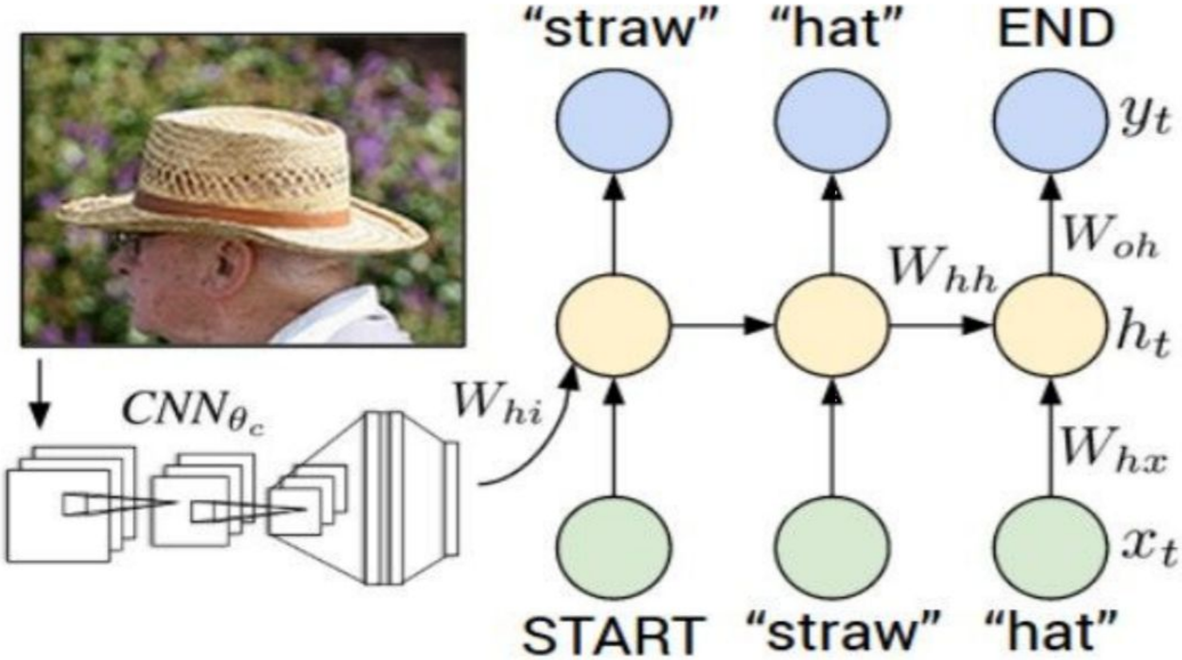
Wednesday, May 10, 1:00 PM - 1:50 PM
– Room 220B

https://gputechconf2017.smarteventscloud.com/connect/sessionDetail.ww?SESSION_ID=109936

Real question is: what to do with generated data or trained GAN model

- Inspect model for biological knowledge
- Supplement lack of samples
- ?

Combined network architectures

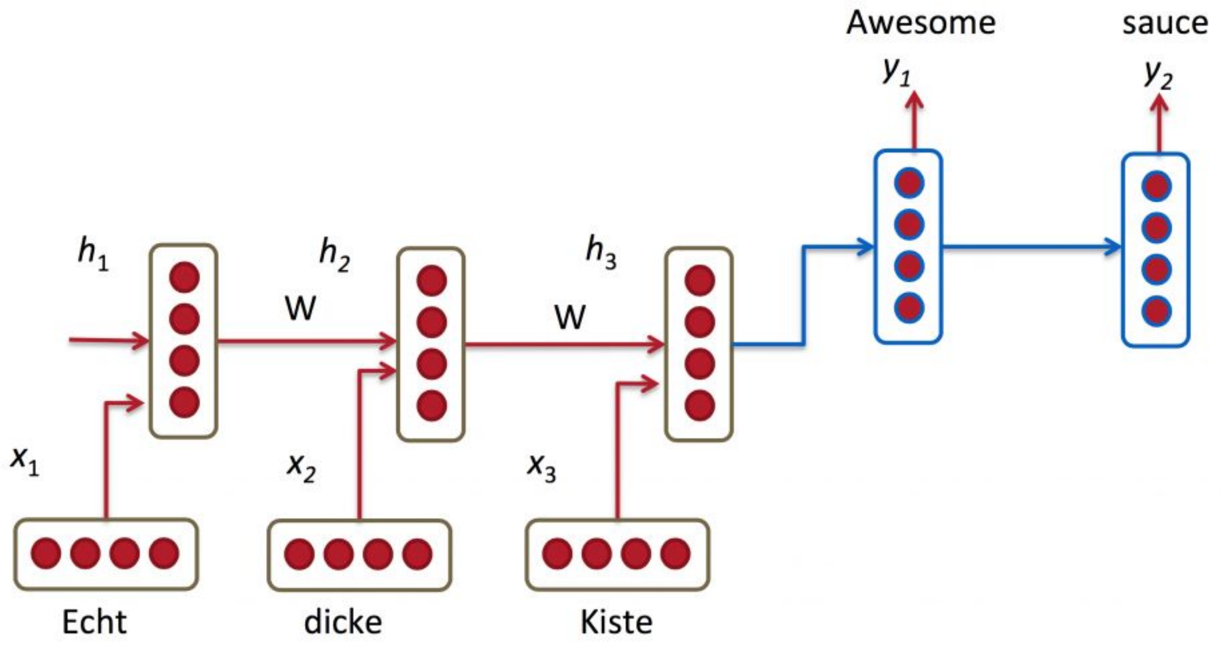


"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

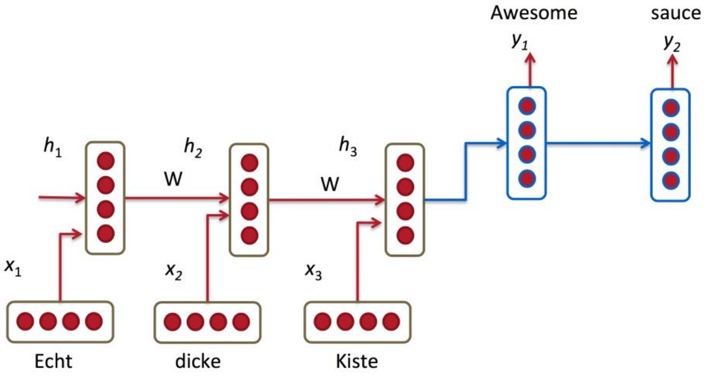
Show, Attend and Tell: Neural Image Caption Generation with Visual Attention
https://arxiv.org/pdf/1502.03044v3.pdf
http://cs231n.stanford.edu/slides/winter1516_lecture10.pdf

Neural translation - encoder/decoder RNNs

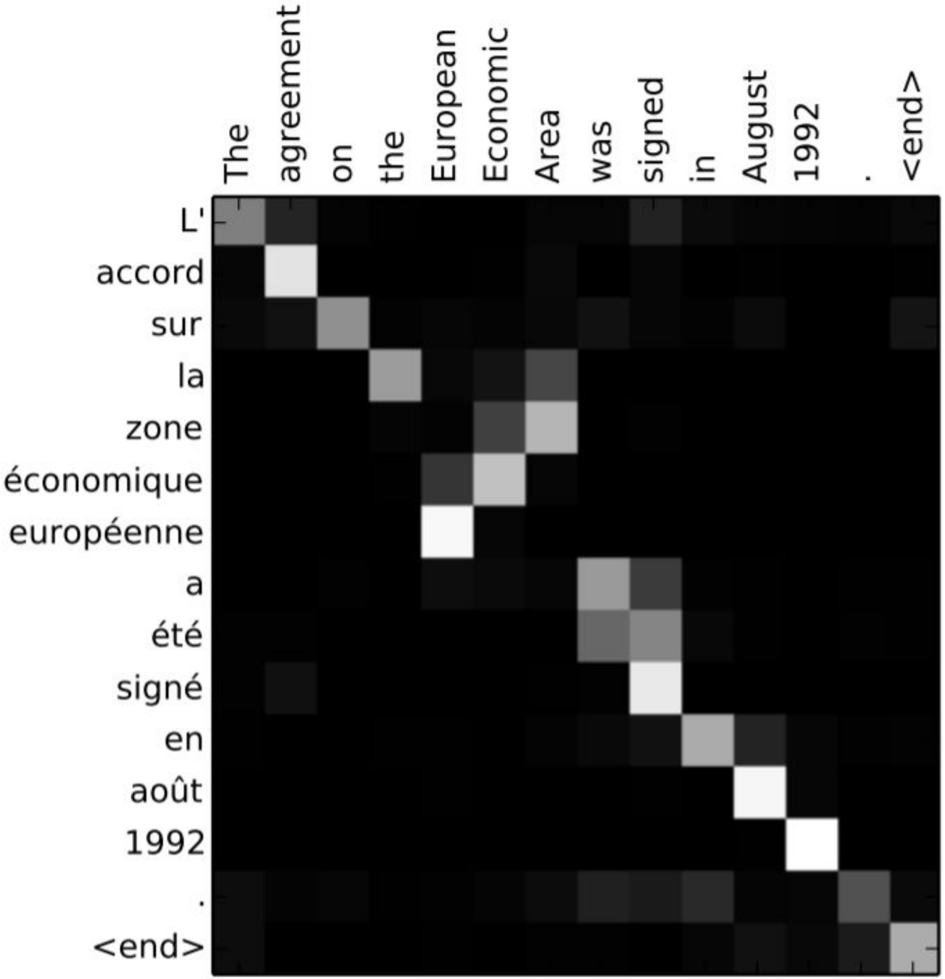


http://www.wildml.com/2016/01/attention-and-memory-in-deep-learning-and-nlp/

Focus over time
Neural Translation

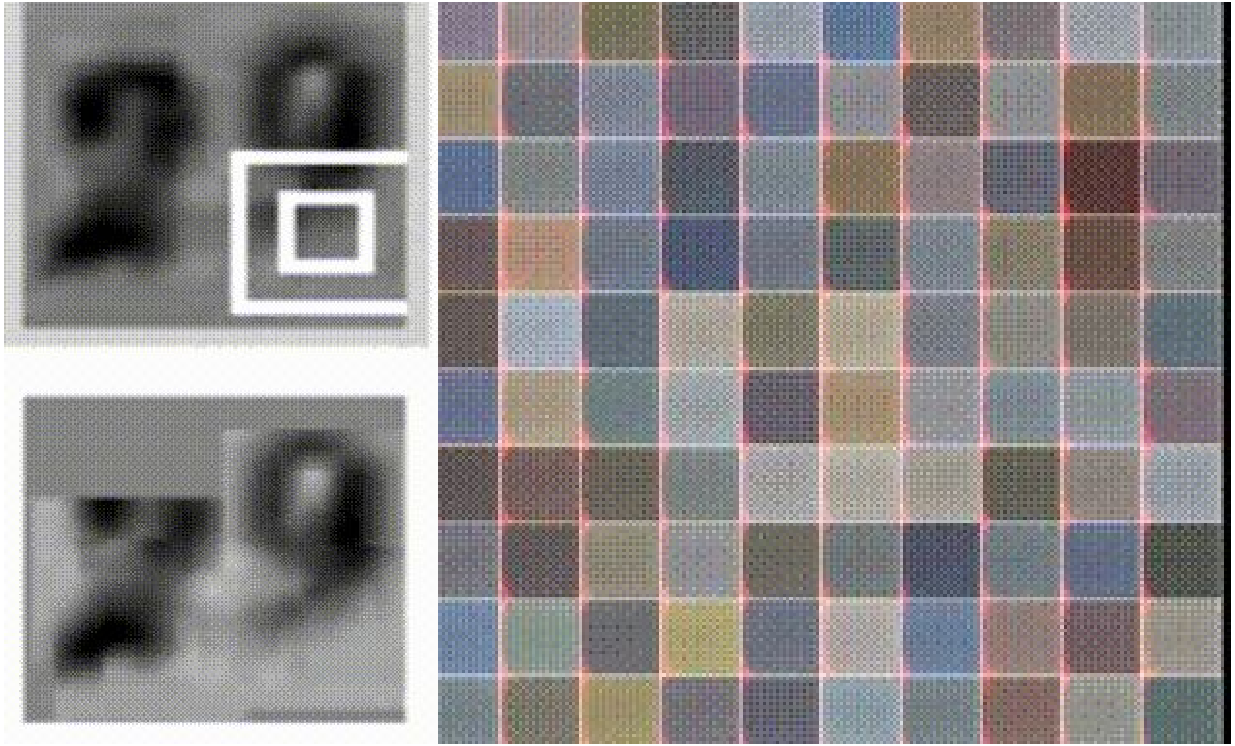


Bahdanau, Cho, Bengio, ICLR 2015



http://www.wildml.com/2016/01/attention-and-memory-in-deep-learning-and-nlp/
https://arxiv.org/pdf/1409.0473.pdf

Attention



http://karpathy.github.io/2015/05/21/rnn-effectiveness/

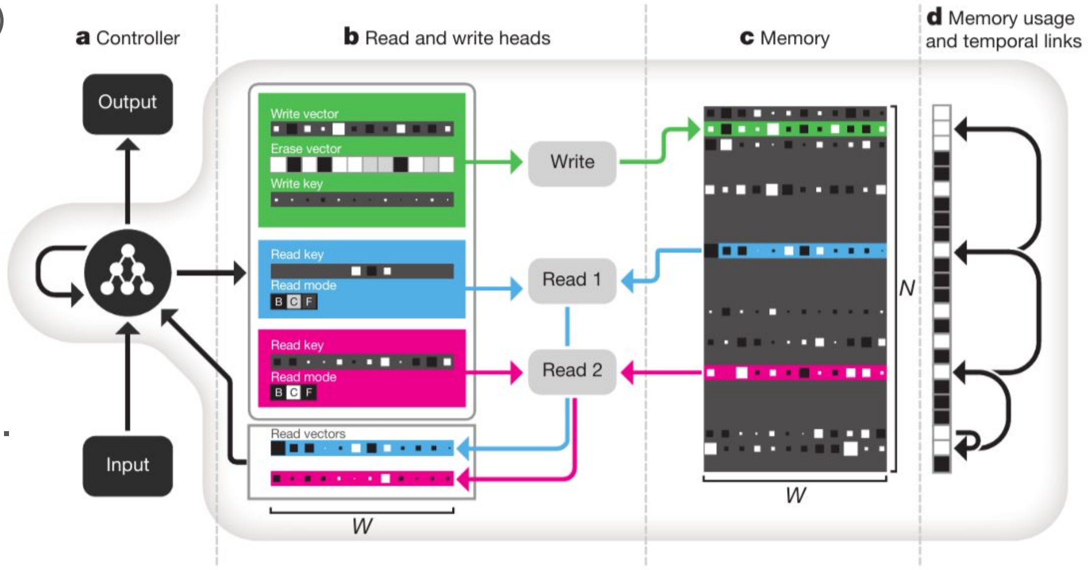
Differential Neural Computers

Modern Computers use:

- Elementary Operations (arithmetic)
- Logic Flow (branching)
- External Memory

Differential Neural Computers mimic these operations in a fully differentiable system, trainable via standard backprop.

- RNN’s
- Attention Models



http://www.nature.com/articles/nature20101.epdf

bAbI dataset
Logic, Reasoning

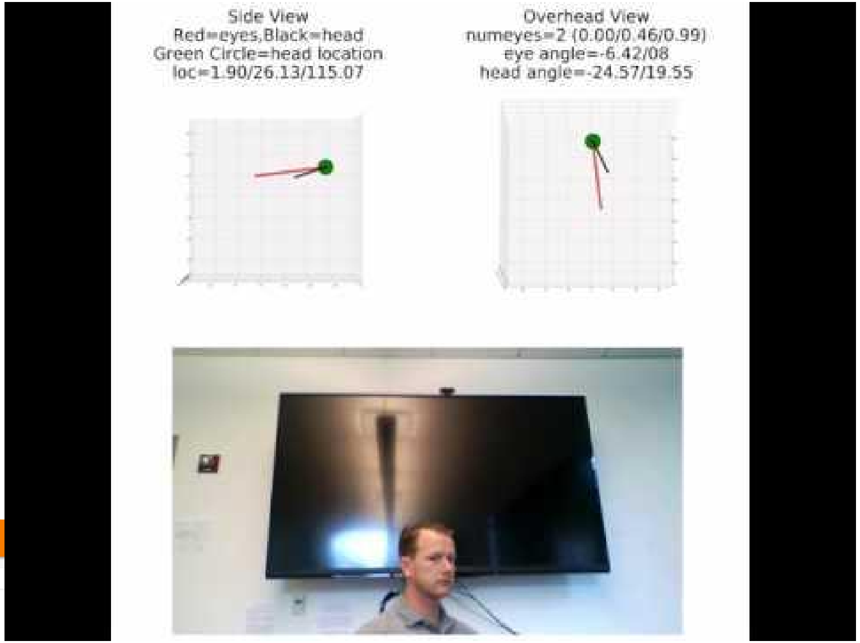
```
mary journeyed to the kitchen . mary moved
to the bedroom . john went back to the
hallway . john picked up the milk there .
what is john carrying ? – john travelled to
the garden . john journeyed to the bedroom
. what is john carrying ? – mary travelled
to the bathroom . john took the apple there
. what is john carrying ? – –
```

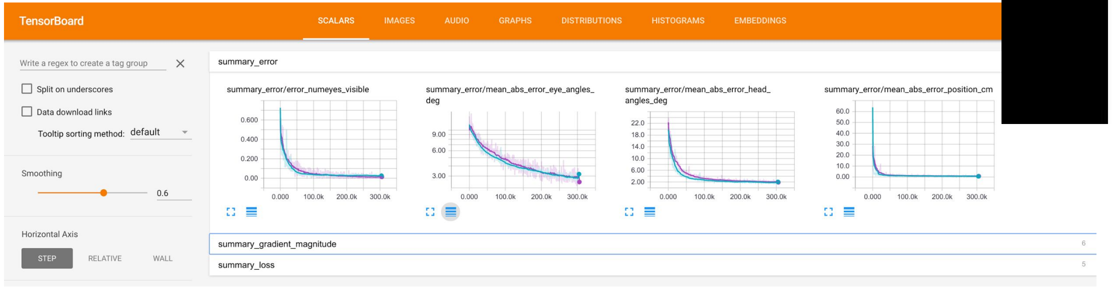
Answers: {milk}, {milk}, {milk apple}

	bAbI Best Results							bAbI Mean Results			
Task	LSTM (Joint)	NTM (Joint)	DNC1 (Joint)	DNC2 (Joint)	MemN2N (Joint) [21]	MemN2N (Single) [21]	DMN (Single) [20]	LSTM	NTM	DNC1	DNC2
1: 1 supporting fact	24.5	31.5	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	28.4 ± 1.5	40.6 ± 6.7	**9.0 ± 12.6**	16.2 ± 13.7
2: 2 supporting facts	53.2	54.5	1.3	0.4	1.0	**0.3**	1.8	56.0 ± 1.5	56.3 ± 1.5	**39.2 ± 20.5**	47.5 ± 17.3
3: 3 supporting facts	48.3	43.9	2.4	**1.8**	6.8	2.1	4.8	51.3 ± 1.4	47.8 ± 1.7	**39.6 ± 16.4**	44.3 ± 14.5
4: 2 argument rels.	0.4	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	0.8 ± 0.5	0.9 ± 0.7	**0.4 ± 0.7**	**0.4 ± 0.3**
5: 3 argument rels.	3.5	0.8	**0.5**	0.8	6.1	0.8	0.7	3.2 ± 0.5	1.9 ± 0.8	**1.5 ± 1.0**	1.9 ± 0.6
6: yes/no questions	11.5	17.1	**0.0**	**0.0**	0.1	0.1	**0.0**	15.2 ± 1.5	18.4 ± 1.6	**6.9 ± 7.5**	11.1 ± 7.1
7: counting	15.0	17.8	**0.2**	0.6	6.6	2.0	3.1	16.4 ± 1.4	19.9 ± 2.5	**9.8 ± 7.0**	15.4 ± 7.1
8: lists/sets	16.5	13.8	**0.1**	0.3	2.7	0.9	3.5	17.7 ± 1.2	18.5 ± 4.9	**5.5 ± 5.9**	10.0 ± 6.6
9: simple negation	10.5	16.4	**0.0**	0.2	**0.0**	0.3	**0.0**	15.4 ± 1.5	17.9 ± 2.0	**7.7 ± 8.3**	11.7 ± 7.4
10: indefinite knowl.	22.9	16.6	0.2	0.2	0.5	**0.0**	**0.0**	28.7 ± 1.7	25.7 ± 7.3	**9.6 ± 11.4**	14.7 ± 10.8
11: basic coreference	6.1	15.2	**0.0**	**0.0**	**0.0**	0.1	0.1	12.2 ± 3.5	24.4 ± 7.0	**3.3 ± 5.7**	7.2 ± 8.1
12: conjunction	3.8	8.9	0.1	**0.0**	0.1	**0.0**	**0.0**	5.4 ± 0.6	21.9 ± 6.6	**5.0 ± 6.3**	10.1 ± 8.1
13: compound coref.	0.5	7.4	**0.0**	0.1	**0.0**	**0.0**	0.2	7.2 ± 2.3	8.2 ± 0.8	**3.1 ± 3.6**	5.5 ± 3.4
14: time reasoning	55.3	24.2	0.3	0.4	**0.0**	0.1	**0.0**	55.9 ± 1.2	44.9 ± 13.0	**11.0 ± 7.5**	15.0 ± 7.4
15: basic deduction	44.7	47.0	**0.0**	**0.0**	0.2	**0.0**	**0.0**	47.0 ± 1.7	46.5 ± 1.6	**27.2 ± 20.1**	40.2 ± 11.1
16: basic induction	52.6	53.6	52.4	55.1	**0.2**	51.8	0.6	**53.3 ± 1.3**	53.8 ± 1.4	53.6 ± 1.9	54.7 ± 1.3
17: positional reas.	39.2	25.5	24.1	**12.0**	41.8	18.6	40.4	34.8 ± 4.1	**29.9 ± 5.2**	32.4 ± 8.0	30.9 ± 10.1
18: size reasoning	4.8	2.2	4.0	**0.8**	8.0	5.3	4.7	5.0 ± 1.4	4.5 ± 1.3	**4.2 ± 1.8**	4.3 ± 2.1
19: path finding	89.5	4.3	**0.1**	3.9	75.7	2.3	65.5	90.9 ± 1.1	86.5 ± 19.4	**64.6 ± 37.4**	75.8 ± 30.4
20: agent motiv.	1.3	1.5	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	1.3 ± 0.4	1.4 ± 0.6	**0.0 ± 0.1**	**0.0 ± 0.0**
Mean Err. (%)	25.2	20.1	4.3	**3.8**	7.5	4.2	6.4	27.3 ± 0.8	28.5 ± 2.9	**16.7 ± 7.6**	20.8 ± 7.1
Failed (err. > 5%)	15	16	**2**	**2**	6	3	**2**	17.1 ± 1.0	17.3 ± 0.7	**11.2 ± 5.4**	14.0 ± 5.0

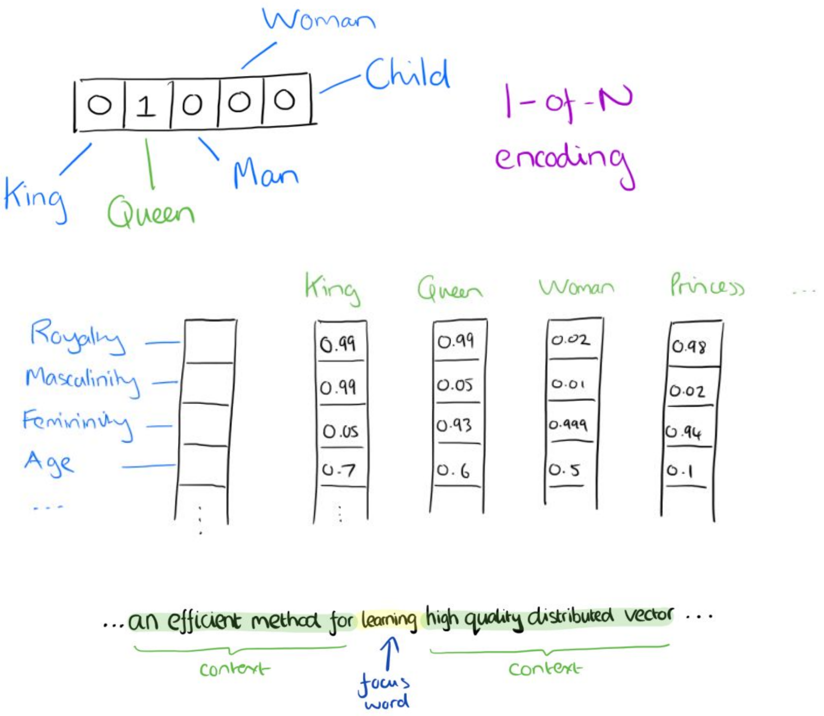
Multi task learning

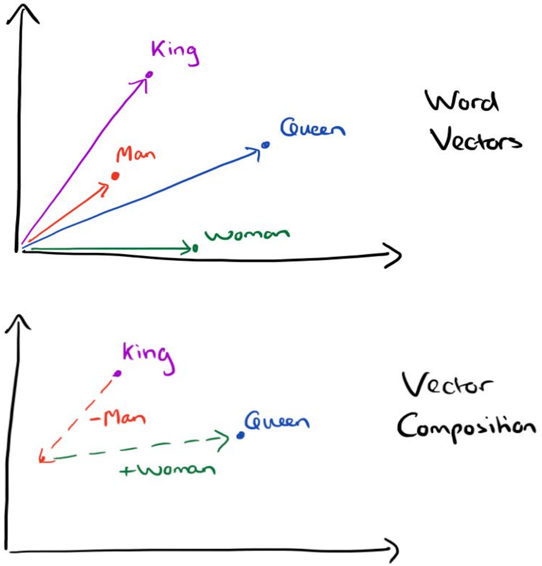
- Convolutional Network
- Only 6,000 data points
- 6M dimensional input (1080x1920)
- No regularization
- No overfitting





Embedding Methods (Example: word2vec)





Query: `King − Man + Woman = ?`

Diagrams by *Adrian Colyer*: https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/

Word2Vec result examples

Newspapers			
New York	New York Times	Baltimore	Baltimore Sun
San Jose	San Jose Mercury News	Cincinnati	Cincinnati Enquirer
NHL Teams			
Boston	Boston Bruins	Montreal	Montreal Canadiens
Phoenix	Phoenix Coyotes	Nashville	Nashville Predators
NBA Teams			
Detroit	Detroit Pistons	Toronto	Toronto Raptors
Oakland	Golden State Warriors	Memphis	Memphis Grizzlies
Airlines			
Austria	Austrian Airlines	Spain	Spainair
Belgium	Brussels Airlines	Greece	Aegean Airlines
Company executives			
Steve Ballmer	Microsoft	Larry Page	Google
Samuel J. Palmisano	IBM	Werner Vogels	Amazon

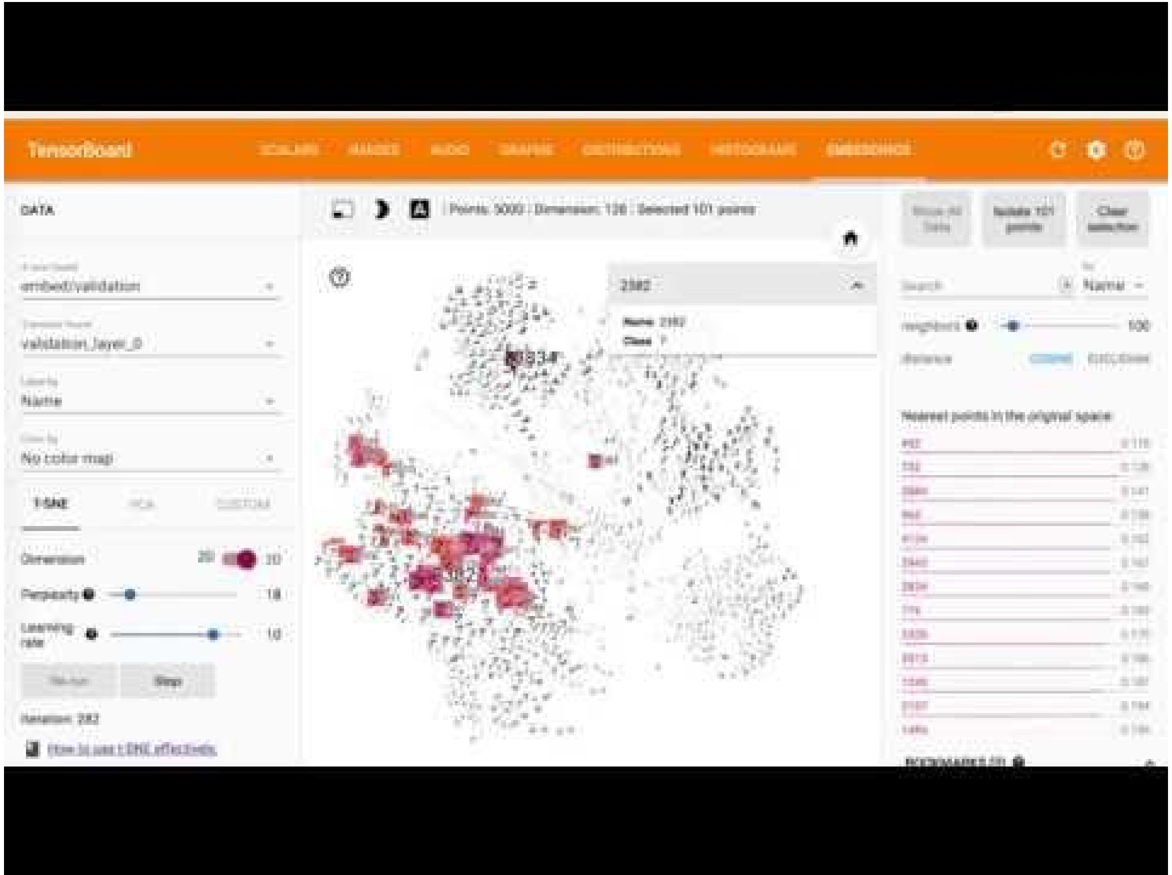
A *is to* B *as* B *is to* ?

Czech + currency	Vietnam + capital	German + airlines	Russian + river	French + actress
koruna	Hanoi	airline Lufthansa	Moscow	Juliette Binoche
Check crown	Ho Chi Minh City	carrier Lufthansa	Volga River	Vanessa Paradis
Polish zolty	Viet Nam	flag carrier Lufthansa	upriver	Charlotte Gainsbourg
CTK	Vietnamese	Lufthansa	Russia	Cecile De

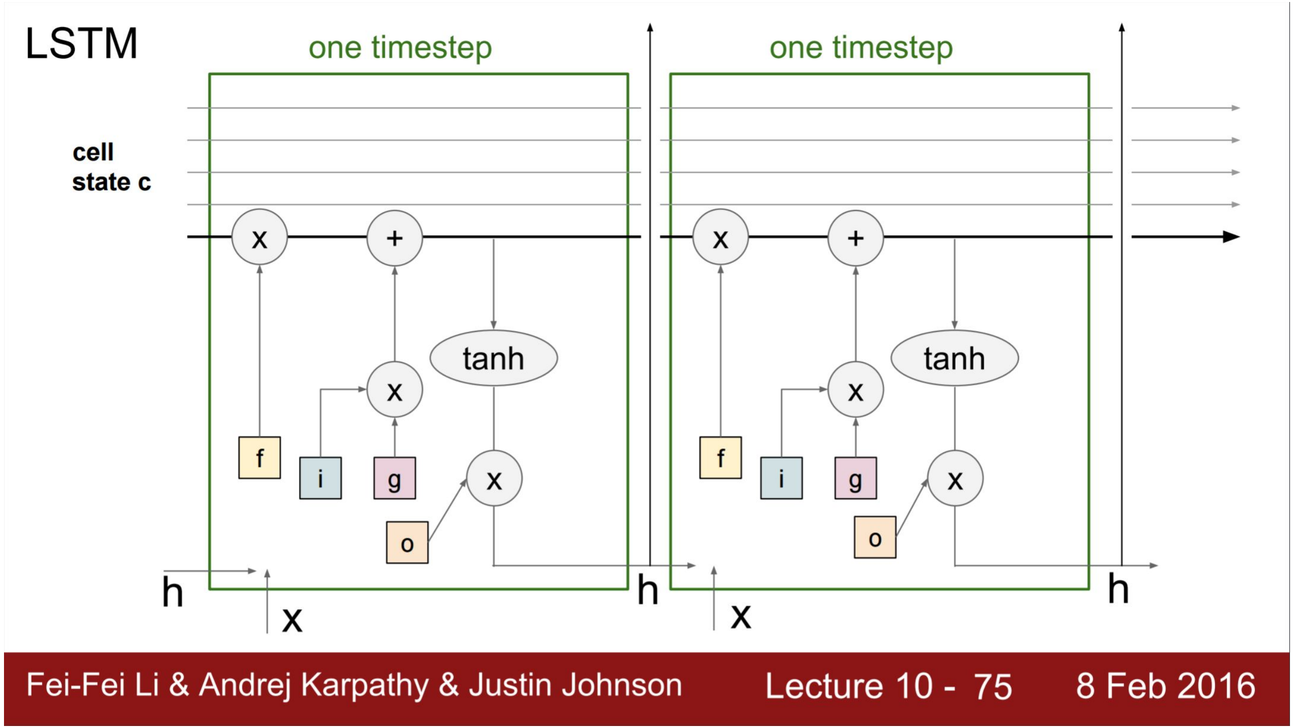
Element wise addition

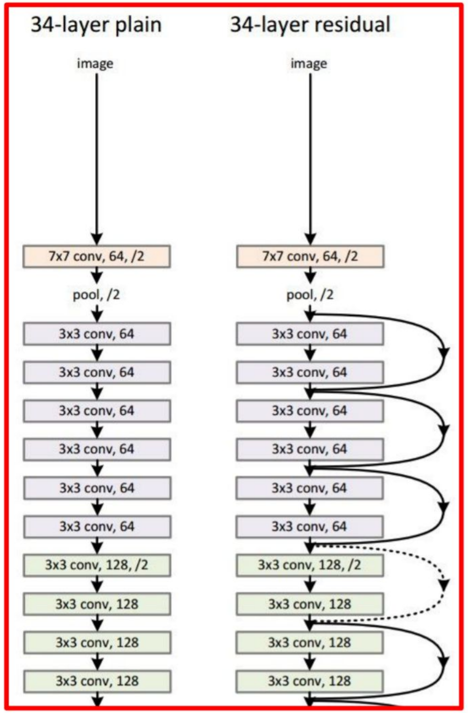
Table 5: Vector compositionality using element-wise addition. Four closest tokens to the sum of two vectors are shown, using the best Skip-gram model.

t-SNE embeddings in TensorBoard



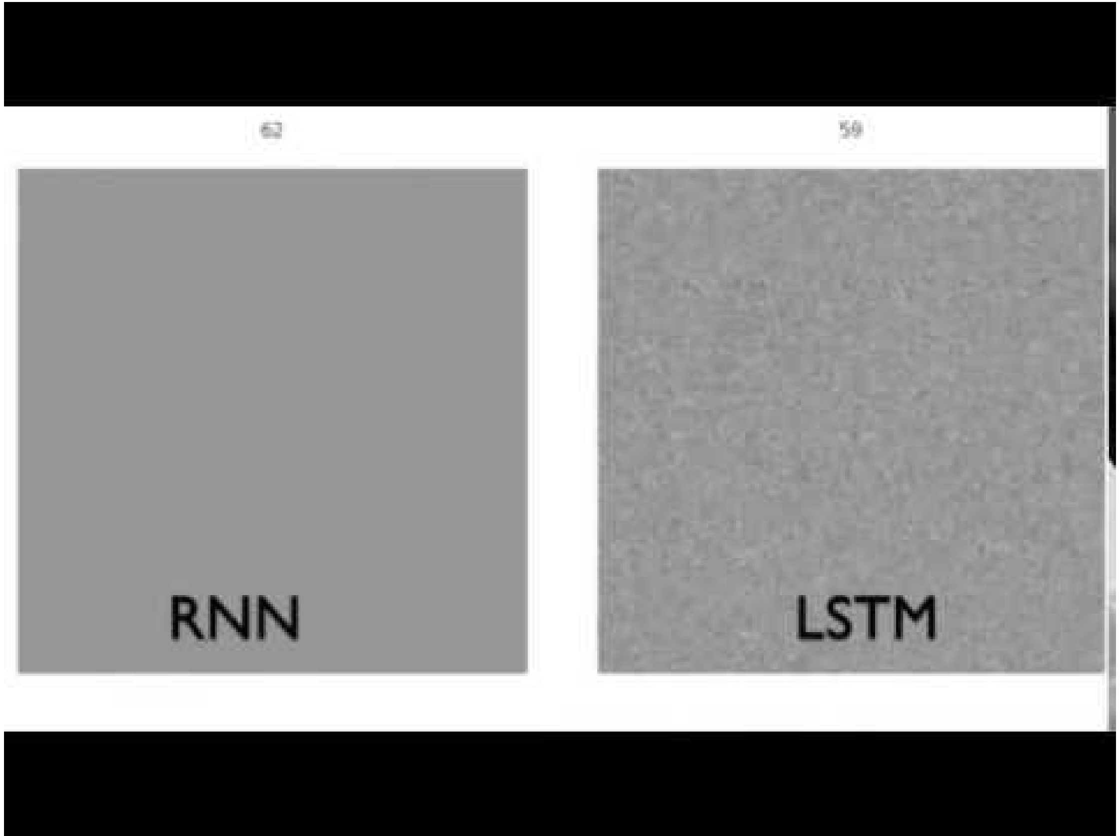
Vanishing gradient solution in LSTM & Residual Nets





* http://cs231n.stanford.edu/slides/winter1516_lecture10.pdf

Vanishing Gradient LSTM vs vanilla RNN



http://imgur.com/gallery/vaNahKE

Deep learning resources

Intro book on neural networks (exceedingly well written, concise):

http://neuralnetworksanddeeplearning.com/

Full course, Stanford's cs231n (video lectures):

http://cs231n.stanford.edu/

The bible of deep learning (complete, well written, in depth):

http://www.deeplearningbook.org/