

Applications and Deep Learning State of the Art

What is Deep Learning?

- Long pipeline of processing operations
- Designed by showing examples
- Example: TUT Age Estimation

<https://youtu.be/Kfe5hKNwrCU>

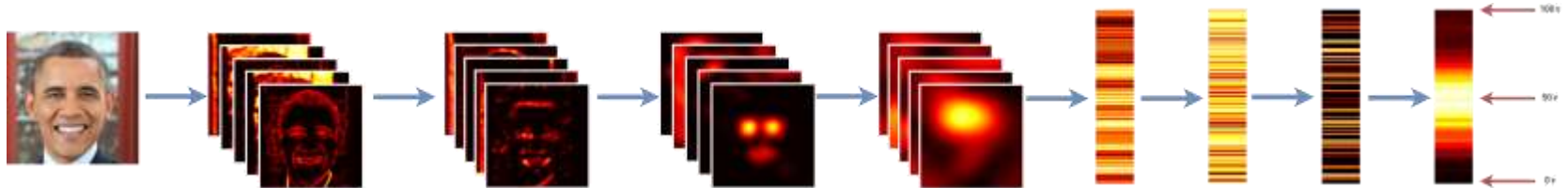


Image Recognition

- Imagenet is the standard benchmark set for image recognition
- Classify 256x256 images into 1000 categories, such as "person", "bike", "cheetah", etc.
- Total 1.2M images
- Many error metrics, including top-5 error: error rate with 5 guesses



Picture from Alex Krizhevsky et al., "ImageNet Classification with Deep Convolutional Neural Networks", 2012

Computer Vision: Case Visy Oy

- Computer vision for logistics since 1994
- License plates (LPR), container codes,...
- How to grow in an environment with heavy competition?
 - Be agile
 - Be innovative
 - Be credible
 - Be customer oriented
 - Be technologically state-of-the-art



Kymmenistätuhansista autoista verot maksamatta – poliisin uusi laite käräytti 74 000 autoa

WUFI 8/12/2015 16:00 Viikkoliikenne 8/12/2015 16:00

Jouko Toppinen PÄÄTOIMITTAJA SAKU/STAFF



What has changes in 20 years?

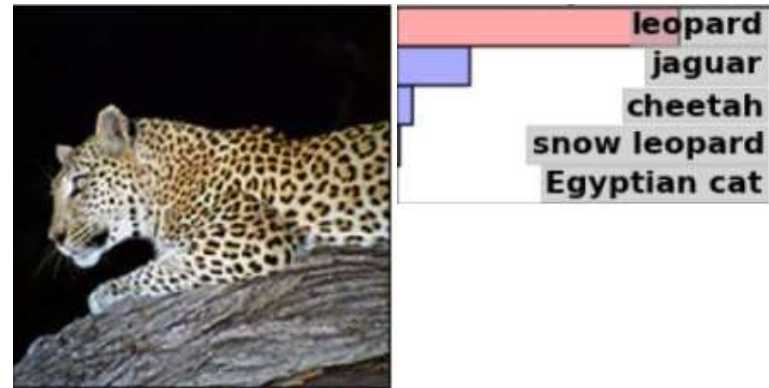
• In 1996:

- Small images (e.g., 10x10)
- Few classes (< 100)
- Small network (< 4 layers)
- Small data (< 50K images)



• In 2016:

- Large images (256x256)
- Many classes (> 1K)
- Deep net (> 100 kerrosta)
- Large data (> 1M)



Net Depth Evolution Since 2012

ILSVRC Image Recognition Task:

- 1.2 million images
- 1 000 categories

(Prior to 2012: 25.7 %)



Team	Year	Place	Error (top-5)	Uses external data
SuperVision	2012	1st	16.4%	no
SuperVision	2012	1st	16.3%	Imagenet 22k
Clarifai	2013	1st	11.7%	no
Clarifai	2013	1st	11.2%	Imagenet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no

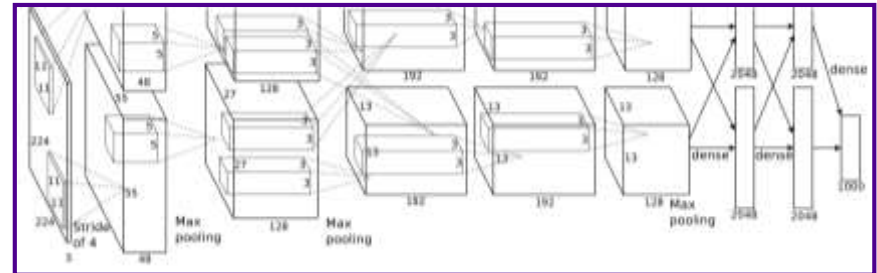
- 2015 winner: MSRA (error 3.57%) → 152 layers
- 2016 winner: Trimps-Soushen (2.99 %) → 152 layers (but many nets)
- 2017 winner: Uni Oxford (2.25 %) → 101 layers (many nets, layers were blocks)

ILSVRC2012

- ILSVRC2012¹ was a game changer
- ConvNets dropped the top-5 error 26.2% → 15.3 %.
- The network is now called *AlexNet* named after the first author (see previous slide).
- Network contains 8 layers (5 convolutional followed by 3 dense); altogether 60M parameters.

The AlexNet

- The architecture is illustrated in the figure.
- The pipeline is divided to two paths (upper & lower) to fit to 3GB of GPU memory available at the time (running on 2 GPU's)
- Introduced many tricks for *data augmentation*
- Left-right flip
- Crop subimages (224x224)



Picture from Alex Krizhevsky et al., "ImageNet Classification with Deep Convolutional Neural Networks", 2012

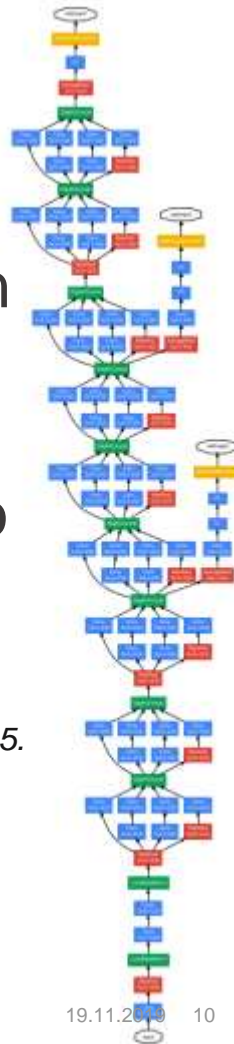
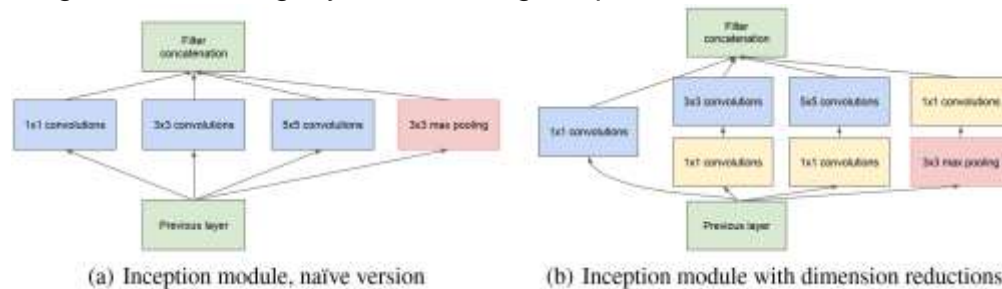
ILSVRC2014

- Since 2012, ConvNets have dominated
- 2014 there were 2 almost equal teams:
 - GoogLeNet Team with 6.66% Top-5 error
 - VGG Team with 7.33% Top-5 error
- In some subchallenges VGG was the winner
- GoogLeNet: 22 layers, only 7M parameters due to fully convolutional structure and clever *inception* architecture
- VGG: 16 layers, 144M parameters

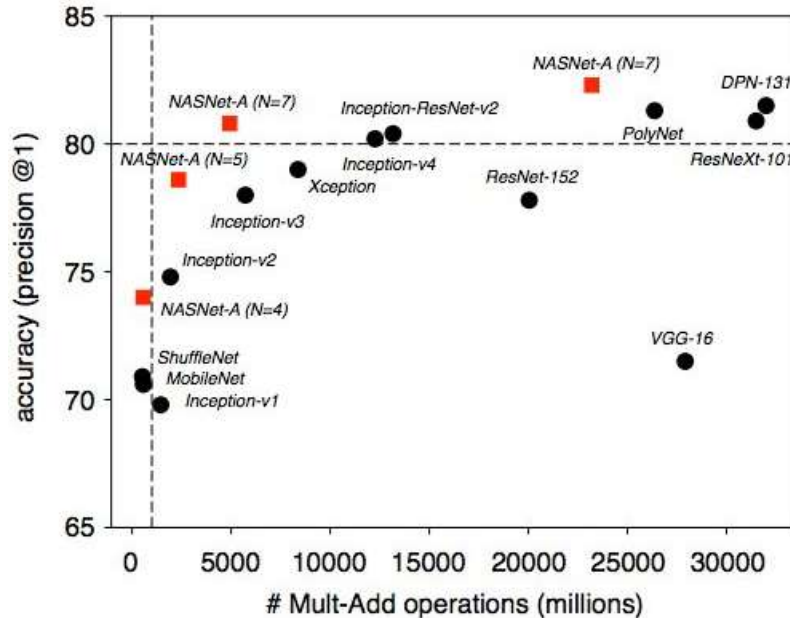
Inception module

- Winner of 2014 ILSVRC (Google) introduced "inception module" in their GoogleNet solution.
- The idea was to apply multiple convolution kernels at each layer, thus reducing the computation compared to then-common 5x5 or 7x7 convolutions.
- Also, the depth was increased by auxiliary losses.

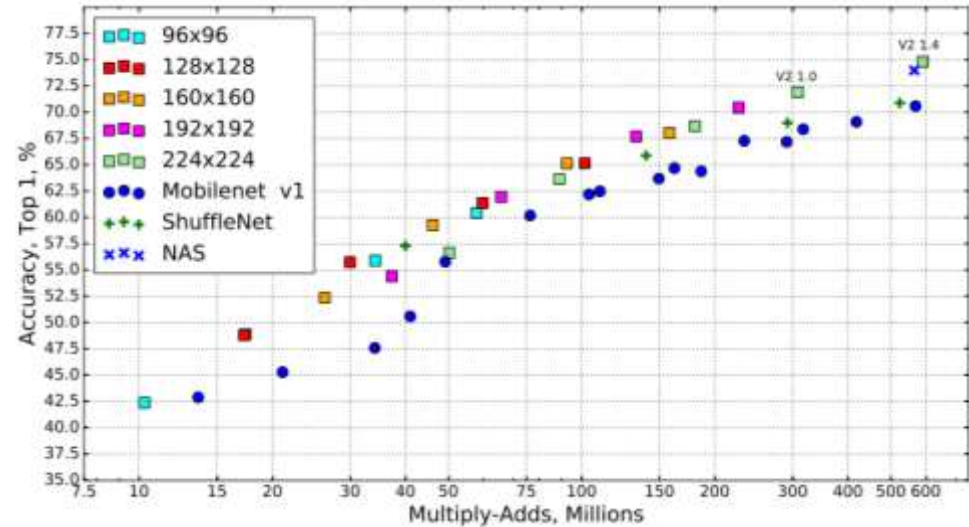
Figures from: Szegedy, et al. "Going deeper with convolutions." CVPR 2015.



Some Famous Networks



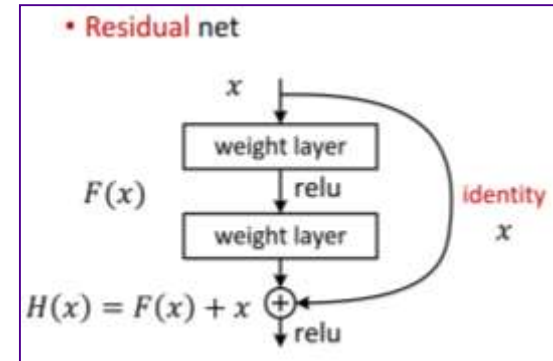
<https://research.googleblog.com/2017/11/automl-for-large-scale-image.html>



Sandler et al., "Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation," Jan. 2018. <https://arxiv.org/abs/1801.04381>

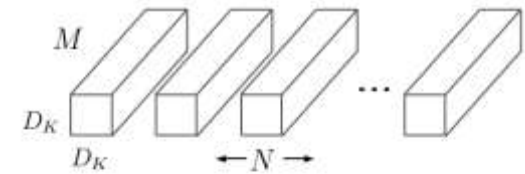
ILSVRC2015

- Winner MSRA (Microsoft Research) with TOP-5 error 3.57 %
- 152 layers! 51M parameters.
- Built from residual blocks (which include the inception trick from previous year)
- Key idea is to add *identity shortcuts*, which make training easier

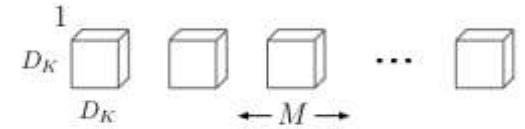


Mobilenets

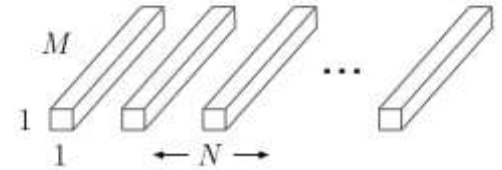
- On the lower end, the common choice is to use **mobilenets**, introduced by Google in 2017.
- Computational load reduced by separable convolutions: each 3x3 conv is replaced by a *depthwise* and *pointwise* convolution.
- Also features a *depth multiplier*, which reduces the channel depth by a factor $\alpha \in \{0.25, 0.5, 0.75, 1.0\}$



(a) Standard Convolution Filters



(b) Depthwise Convolutional Filters



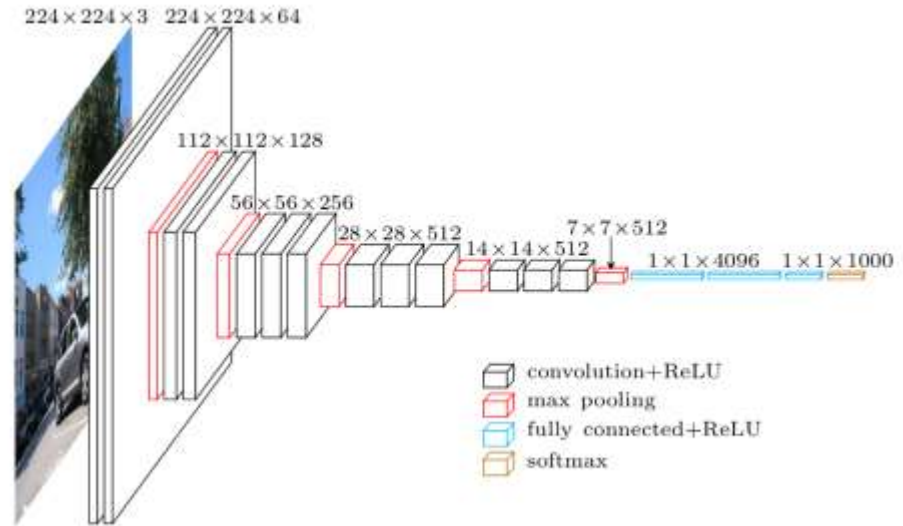
(c) 1×1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

Table 6. MobileNet Width Multiplier

Width Multiplier	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
0.75 MobileNet-224	68.4%	325	2.6
0.5 MobileNet-224	63.7%	149	1.3
0.25 MobileNet-224	50.6%	41	0.5

Pretraining

- With small data, people often initialize the net with a *pretrained* network.
- This may be one of the imagenet winners; VGG16, ResNet, ...
- See `keras.applications` for some of these.



VGG16 network

source: <https://www.cs.toronto.edu/~frossard/post/vgg16/>

Example: Cats vs. Dogs

- Let's study the effect of pretraining with classical image recognition task: learn to classify images to **cats** and **dogs**.
- We use the *Oxford Cats and Dogs* dataset.
- Subset of 3687 images of the full dataset (1189 cats; 2498 dogs) for which the ground truth location of the animal's head is available.



Network 1: Design and Train from Scratch

```
# Initialize the model
model = Sequential()

shape = (64, 64, 3)

# Add six convolutional layers. Maxpool after every second convolution.
model.add(Conv2D(filters=32, kernel_size=3, padding="same", activation="relu",
input_shape=shape))
model.add(Conv2D(filters=32, kernel_size=3, padding="same", activation="relu"))
model.add(MaxPooling2D(2, 2)) # Shrink feature maps to 32x32

model.add(Conv2D(filters=48, kernel_size=3, padding="same", activation="relu"))
model.add(Conv2D(filters=48, kernel_size=3, padding="same", activation="relu"))
model.add(MaxPooling2D(2, 2)) # Shrink feature maps to 16x16

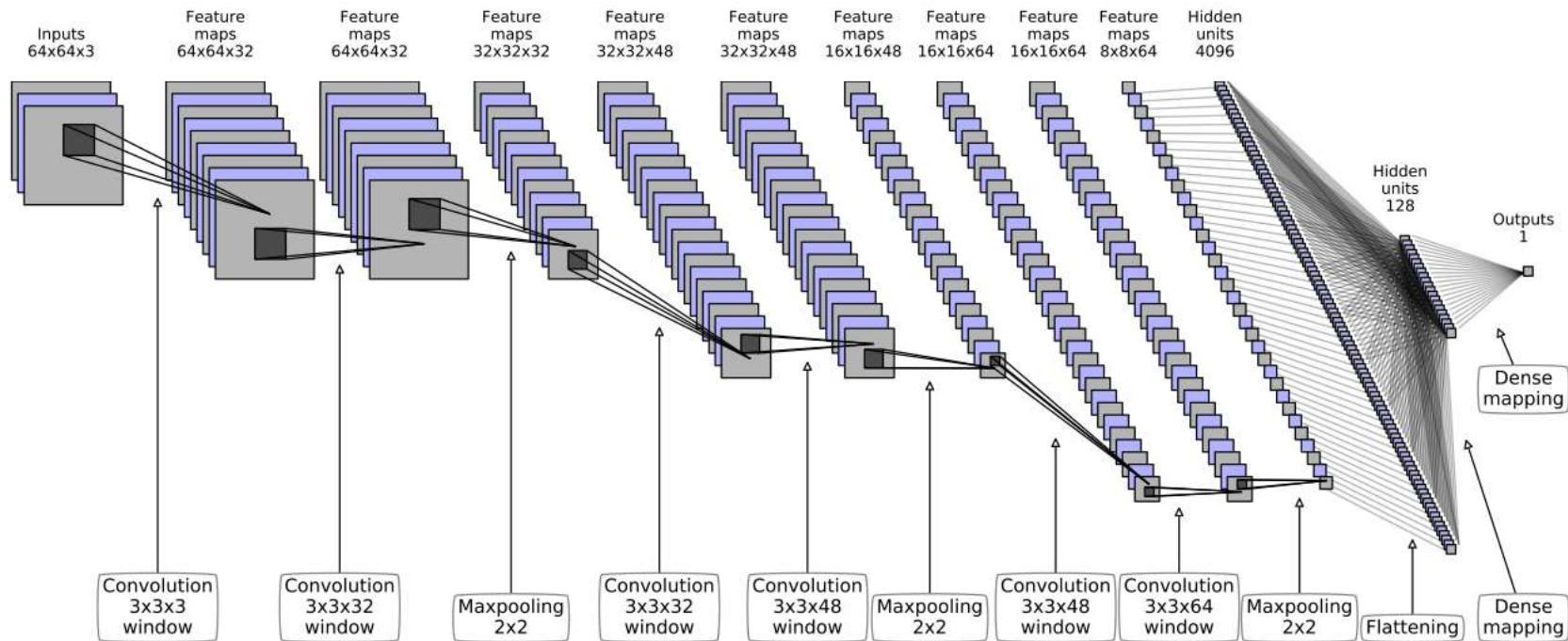
model.add(Conv2D(filters=64, kernel_size=3, padding="same", activation="relu"))
model.add(Conv2D(filters=64, kernel_size=3, padding="same", activation="relu"))
model.add(MaxPooling2D(2, 2)) # Shrink feature maps to 8x8

# Vectorize the 8x8x64 representation to 4096x1 vector
model.add(Flatten())

# Add a dense layer with 128 nodes
model.add(Dense(128, activation="relu"))
model.add(Dropout(0.5))

# Finally, the output layer has 1 output with logistic sigmoid nonlinearity
model.add(Dense(1, activation="sigmoid"))
```


Network 1: Design and Train from Scratch



Network 2: Start from a Pretrained Network

```
# Import the network container and the three types of layers
from keras.applications.vgg16 import VGG16
from keras.models import Model
from keras.layers import Dense

# Initialize the VGG16 network. Omit the dense layers on top.
base_model = VGG16(include_top = False, weights = "imagenet",
input_shape = (64, 64, 3))

# We use the functional API, and grab the VGG16 output here:
w = base_model.output

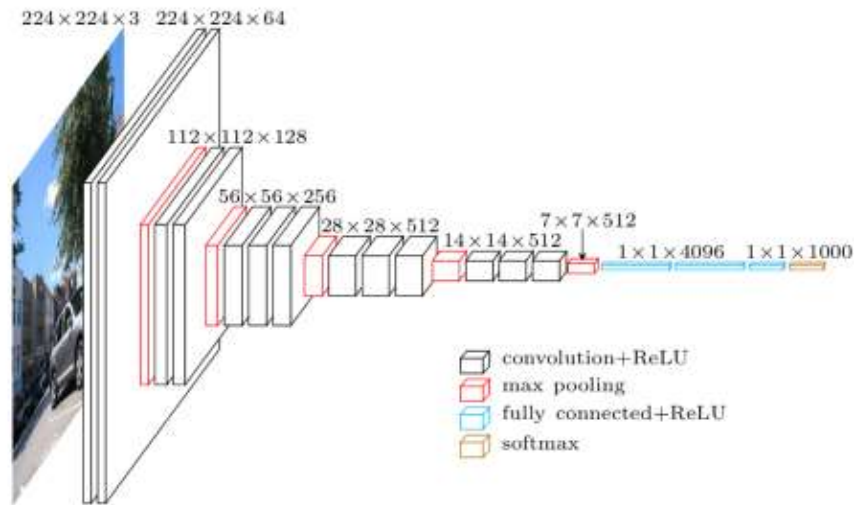
# Now we can perform operations on w. First flatten it to 4096-dim vector:
w = Flatten()(w)

# Add dense Layer:
w = Dense(128, activation = "relu")(w)

# Add output Layer:
output = Dense(1, activation = "sigmoid")(w)

# Prepare the full model from input to output:
model = Model(inputs = [base_model.input], outputs = [output])

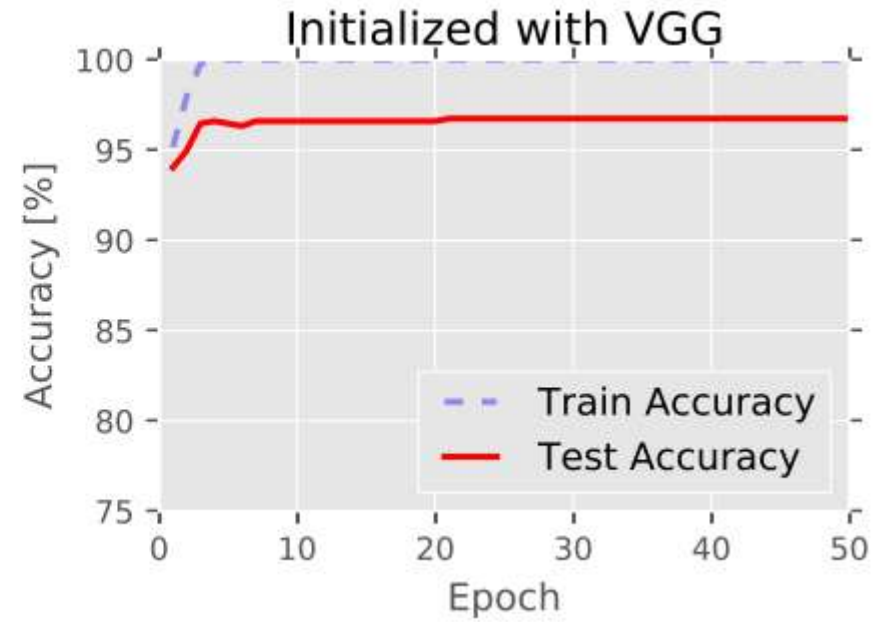
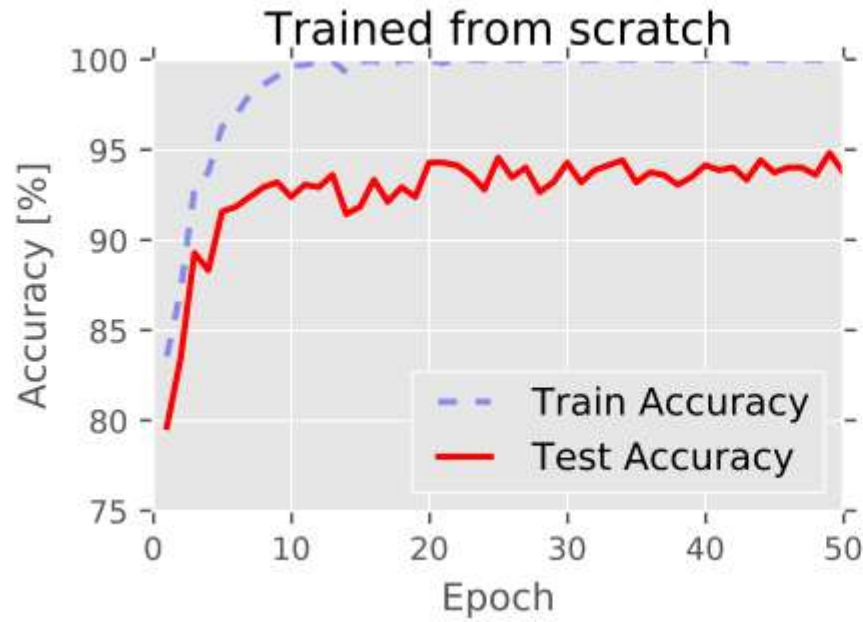
# Also set the last Conv block (3 Layers) as trainable.
# There are four layers above this block, so our indices
# start at -5 (i.e., last minus five):
model.layers[-5].trainable = True
model.layers[-6].trainable = True
model.layers[-7].trainable = True
```



VGG16 network

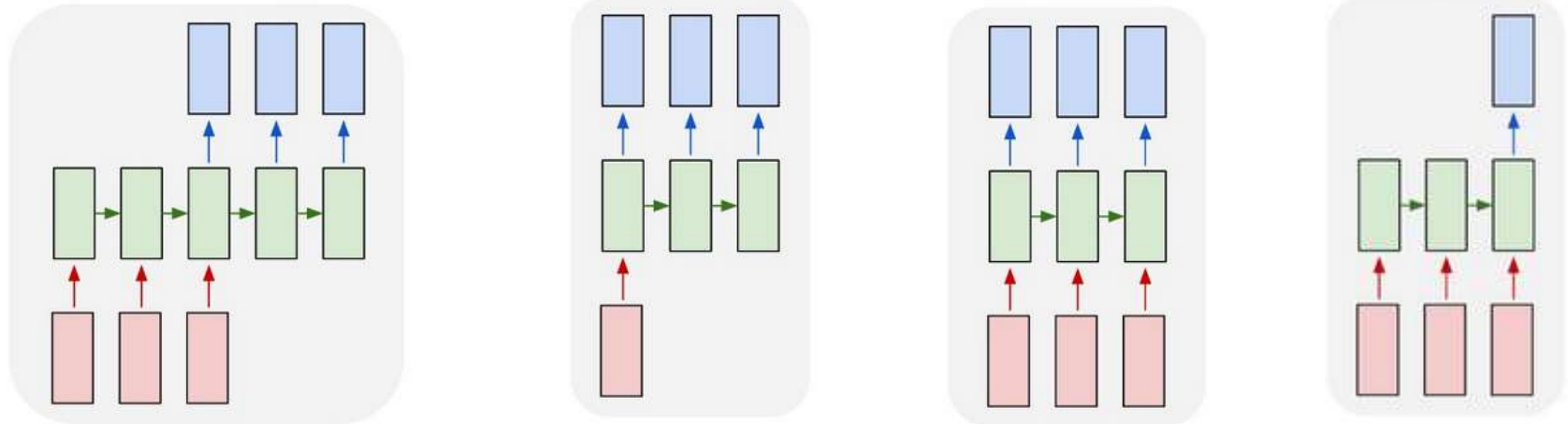
source: <https://www.cs.toronto.edu/~frossard/post/vgg16/>

Results



Recurrent Networks

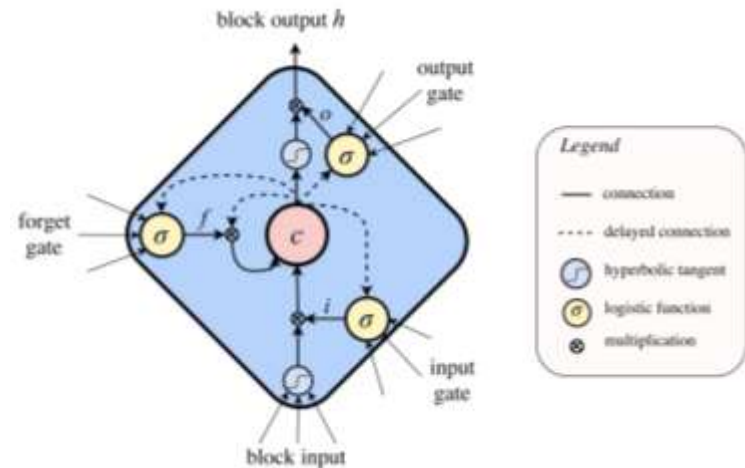
- Recurrent networks process sequences of arbitrary length; e.g.,
 - Sequence → sequence
 - Image → sequence
 - Sequence → class ID



Picture from <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>

Recurrent Networks

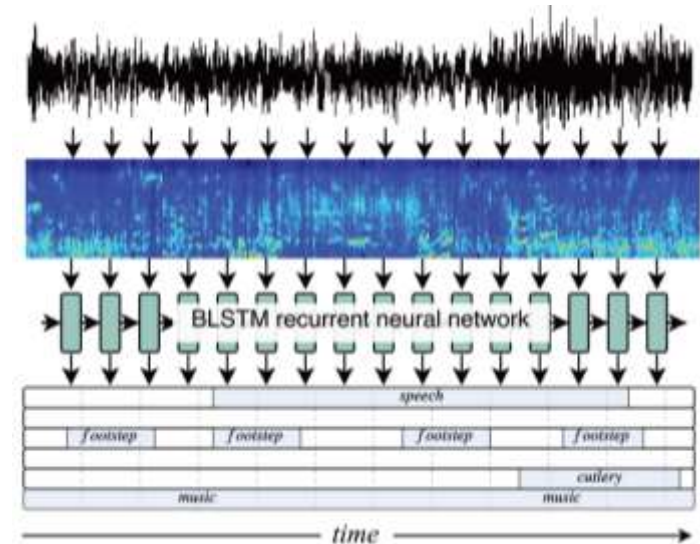
- Recurrent net consist of special nodes that remember past states.
- Each node receives 2 inputs: the data and the previous state.
- Keras implements *SimpleRNN*, *LSTM* and *GRU* layers.
- Most popular recurrent node type is *Long Short Term Memory* (LSTM) node.
- LSTM includes also *gates*, which can turn on/off the history and a few additional inputs.



Picture from G. Parascandolo M.Sc. Thesis, 2015.
<http://urn.fi/URN:NBN:fi:tyy-201511241773>

Recurrent Networks

- An example of use is from our recent paper.
- We detect acoustic events within 61 categories.
- LSTM is particularly effective because it remembers the past events (or the context).
- In this case we used a *bidirectional* LSTM, which remembers also the future.
- BLSTM gives slight improvement over LSTM.



LSTM in Keras

- LSTM layers can be added to the model like any other layer type.
- This is an example for natural language modeling: *Can the network predict next symbol from the previous ones?*
- Accuracy is greatly improved from N-Gram etc.

```
model = Sequential()

model.add(LSTM(512, return_sequences=True,
              input_shape=(maxlen, len(symbols))))
model.add(Dropout(0.2))

model.add(LSTM(512, return_sequences=False))
model.add(Dropout(0.2))

model.add(Dense(len(symbols)))
model.add(Activation('softmax'))

model.compile(loss='categorical_crossentropy', optimizer='rmsprop')
```

Text Modeling

- The input to LSTM should be a sequence of vectors.
- For text modeling, we represent the symbols as binary vectors.

```
from sklearn import preprocessing

lb = preprocessing.LabelBinarizer()
symbol_list = list("hello world")
lb.fit(symbol_list)
binary_table = lb.transform(symbol_list)
```

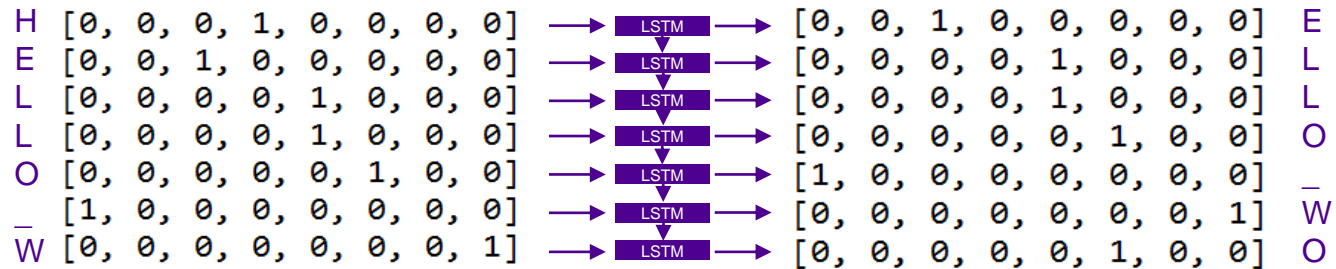


Time ↓

	<u>_</u>	<i>d</i>	<i>e</i>	<i>h</i>	<i>l</i>	<i>o</i>	<i>r</i>	<i>w</i>
array([[0	0	0	1	0	0	0	0]
	0	0	1	0	0	0	0	0]
	0	0	0	0	1	0	0	0]
	0	0	0	0	1	0	0	0]
	0	0	0	0	0	1	0	0]
	1	0	0	0	0	0	0	0]
	0	0	0	0	0	0	0	1]
	0	0	0	0	0	1	0	0]
	0	0	0	0	0	0	1	0]
	0	0	0	0	1	0	0	0]
	0	1	0	0	0	0	0	0]]

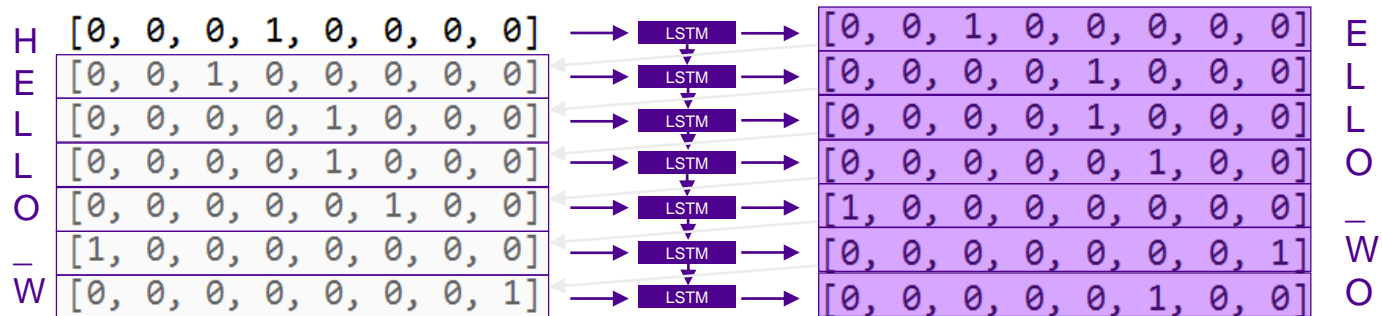
Text Modeling

- The prediction target for the LSTM net is simply the input delayed by one step.
- For example: we have shown the net these symbols: ['h', 'e', 'l', 'l', 'o', '_', 'w']
- Then the network should predict 'o'.



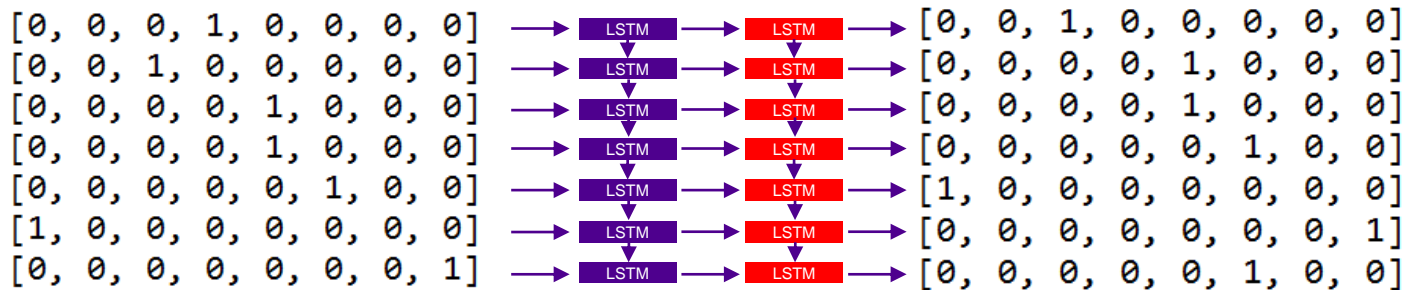
Text Modeling

- Trained LSTM can be used as a text generator.
- Show the first character, and set the predicted symbol as the next input.
- Randomize among the top scoring symbols to avoid static loops.



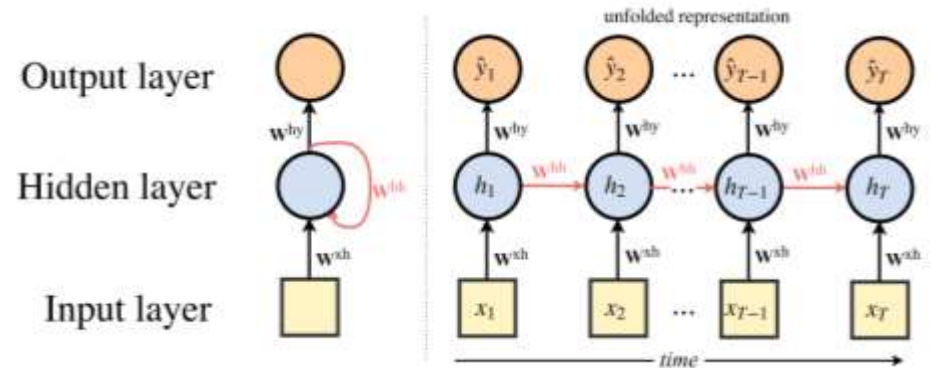
Many LSTM Layers

- A straightforward extension of LSTM is to use it in multiple layers (typically less than 5).
- Below is an example of two layered LSTM.
- Note: Each blue block is exactly the same with, *e.g.*, 512 LSTM nodes. So is each red block.



LSTM Training

- LSTM net can be viewed as a very deep non-recurrent network.
- The LSTM net can be *unfolded* in time over a sequence of time steps.
- After unfolding, the normal gradient based learning rules apply.



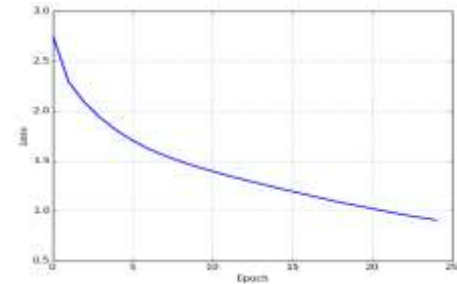
Text Modeling Experiment

- Keras includes an example script:
https://github.com/fchollet/keras/blob/master/examples/lstm_text_generation.py
- Train a 2-layer LSTM (512 nodes each) by showing Nietzsche texts.
- A sequence of 600901 characters consisting of 59 symbols (uppercase, lowercase, special characters).

```
SUPPOSING that Truth is a woman--what then? Is there not ground  
for suspecting that all philosophers, in so far as they have been  
dogmatists, have failed to understand women--that the terrible  
seriousness and clumsy importunity with which they have usually paid  
their addresses to Truth, have been unskilled and unseemly methods for  
winning a woman? Certainly she has never allowed herself to be won
```

Text Modeling Experiment

- The training runs for a few hours on a Nvidia high end GPU (Tesla K40m).
- At start, the net knows only a few words, but picks up the vocabulary rather soon.



it is the sere the the the the and
the the and of the hos an the the
and and the the the the the an
the the the the and the the ant an
the and on the the the he the the
he hor an the the hore the the the
he the he ans ante an the anle the
and and of the the hor and the the
the the he the the the the the
the he the the the and the the the
the the the and of an the the he
the the the the the the he

Epoch 1

artists if they happored and for
the concerced and man actored of
shere all their accorition of the
world the belirition and in the all
of the prose qoominity and in no
berigation of the expores in a
dondicted and for
the forter prosoment that a
condicted and of the menters of the
soul of the dost and the for the

Epoch 3

manifold was not the little have a
strong and contrary, his can be
true to be a great need in the will
to prove and consequence
in short, something hably on the
development of the intellectual and
truth, and consequently, a little
truth and possible all the higher
things than the mastering sense of
the
servant of the enjigh of the sense
of the serve (and who has the goal
it of this fantastic and the di

Epoch 25

Text Modeling Experiment

- Let's do the same thing for Finnish text: All discussions from Suomi24 forum are released for public.
- The message is nonsense, but syntax close to correct: A foreigner can not tell the difference.

kusta siin koista siin kuusta siin
kuiken kaisin kuukan kuinan koikan
ja kainan kuiten kain tuinen kuinan
kuisen siin kuinin siin kutta sitä
koista siin taikaa tuiten sain
koina siin kaikan kuitan eli siin
tiinen suin tuiten siin siitä
kuikaa siitä kuin tuin kankaa kuin
vaitan kuinan tuinen kiinin kaitaa
kaikaan kuinen kuka siinen kun
kuina kutta ja taisin kain
kaikaisin koin kaikon kainan kuina

Epoch 1

niin se vaikka en ole ole kokemista
koko on talletuksen jos on
tarvitalle vaan muutansa tulee
voimattaa koko paljon ja henkin
alkoita ja kanvattaa ovat joskaan
hänen taivalliset kokotalle
toiminetto en ole maanaan.

suukaan tule vielä koitaan saa
varhan haluaa elämään se jotain
toisesta olen työnyt tulee en ole
vaikka sanon tapahtamisen raukan

Epoch 4

mitään toisten on kokemusta kuin tehdä
sinun vielä kerran vaihtaa kun olen
kokeillut maan kanssa. ja sitten tulisi
halua kaikki kaupat talletukset.

- paras grafiikka peleissä ja ulkoasussa

ensimmäinen bonus: 10 ilmaiskierrosta
peliin liikkun kunnon tuomittaa kun ei
ole valita yksi alla on
kerrottu sitä miten saattaa minun kanssa.
samoin suustaa kokonaan ja painan si

Epoch 44

Fake text

- February 2019: "Dangerous"

Our model
predicts
about the
the train
instead
with, as

SYSTEM PROMPT
(HUMAN-WRITTEN)

A train
Cincinnati

MODEL
COMPLETION
(MACHINE-
WRITTEN, FIRST
TRY)

The in
Coving

In an
working

"The tl
consequ
economy
statement
happen

SYSTEM PROMPT
(HUMAN-WRITTEN)

Recycling is good for the world.

NO! YOU COULD NOT BE MORE WRONG!!

MODEL
COMPLETION
(MACHINE-
WRITTEN, 25
TRIES)

Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I'm not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. Recycling is bad for our economy. It increases the cost of a product, and in turn, the price of everything that is made with that product. Recycling is not good for our nation. We pay a tremendous price for the privilege of having the world's most advanced and efficient recycling system. Recycling is a huge, colossal waste of time, energy, money, and resources. And THAT is why we need to get back to basics and get back to basics in our recycling efforts. One of the best ways to start is to look at the process of creating a paper product. When you make a paper product, it is basically a long chain of materials. Everything from the raw materials (wood, cardboard, paper, etc.), to the reagents (dyes,

Suomi24 generator

- We train the OpenAI model with Suomi24 corpus.
- After 300 iterations, the text resembles Finnish.

```
"C-3a": Tuhutunutkimman yhtään sökivaltii, että on erikkin kirjoituksen kun  
noin ihmisiä yhtään söksyt lomassa.  
Hintaa tulla olen kirjoituksen läpi yms.
```

```
"kurinkoa": kirjoituksen kurinkoa. Itse vai kirjoitettavuotia kirjoituksen  
kirjoitettavuotia. Täyttävä on yhdröttäväksyölkükun vai kirjoituksen kirjo  
ituksen kirjoituksen kirjoituksen kirjoituksen kirjoituksen kirjoituksen ki  
rjoituksen kirjoituksen kirjoituksen kirjoituksen kirjoituksen kirjoituksen  
kirjoituksen kirjoituksen kirjoituksen kirjoituksen kirjoituksen kirjoituk  
sen kirjoituksen kirjoituksen kirjoituksen kirjoituksen kirjoituksen kirjoi  
tuksen kirjoituksen kirjoituksen kirjoituksen kirjoituksen kirjoituksen kir  
joituksen kirjoituksen kirjoituksen kirjoituksen kirjoituksen kirjoituksen  
kirjoituksen kirjoituksen kirjoituksen kirjjaa toivoa ihmisessä.
```

```
"B-tjlu4": Ihmiset yhtään söksyölkärssään ikinä minkäänä kirjoituksesta oli  
kirjoituksesta ikinät edes kirjoituksesta kirjoituksesta. Varmaan kohden l  
isäksi kuin yhtään söksyölkärssä todisteita ja nyt ikinää.
```

After 10000 iterations

```
"älypuuhuva": Näin menee aivan turhaan.  
Ei tällainen tai ajatustainen. Jalkoista ja jos sitä lasketaankin, niin sil  
  
"Lapset ärsyttää": VITUN VÄELYN MUIMAA  
  
"kävä": Sama, ite oon käynny hiuksen takia  
  
"Oon tässä": Olisin ihan kiva saada uusia öljypukkiia makkulauhat ja se tar  
tyyppi joka pystyi makkulauhan. Lähinnä ku on raska.  
http://www.suomi24.fi/  
http://www.suomi24.fi/kotimaa/17126955.html  
  
"nauttii": No, täällä on turhan kuvaaja maksaa ja lainkaan. Itse oon jo suu  
rvee jäikin  
  
"Tää mättö": No joka oikais  
täällä on sisaren korkeammat pojat, joka pystyy sisarenkahinnassa. Joku on  
isaren oirehtimisessä. Onneks näiden pitää lähteä hampaat hiihtoon elämänva  
  
"Oon tottu": Jos olisi syvää näin jalat suunnitella jollain kuukautta, jos  
kauhajaan, voisin pätee poikkiin. Kuvaaja maksaa vieläkin uusiksi, ilmeises  
ä poikkiin, kuin se uusi aloitus johdattaa, ja sen takia, minä en. Ompa its  
la pimeitä, mutta en pääse poikkiin.
```

After 380000 iterations

"Keskivaarassa asuva": Jokaiselle vois löytää oman asunnon. Löytyy osa muilta sivuilta luonnosta, ja joskus voi löytää saa löytyä netistä ja käytettyjä, etenkin tietokonettakin saatavaa aineistoa. Esim. keskustelu sivuilla <http://www.kristillisdemokraatit.se?e=>

"R. Sitia": eivät ole mitään "kristillisdemokraateja" tai jos on, niin kai sinne haluaa löytää tämmöistä suomalaista, joka on muutakin kuin omassa keskustelussaan. Ja luonnosta eivät ole niin väliksi, että kyseessä olisi siirtolaispappi Väinämö Voitosta. Hintatasolla on vaikea sanoa, jos pinnisarvi on kohdallaan, niin toki pitää tätäkin palstaa tutkia ennen kuin katselee kirjoituksia. Suurin osa kirjoittajista on tarkoitettu sattumalta suomalaisten elämänmuodostelmiin. Onko se muuten näin?

"Diana*": sinänsä pahaa, että ihmisiä on kirkon alaisuudessa pikku hiljaa viety ja lopussa kaikilla on jonkinlainen yhteistoiminta. Olen tavannut tällaisen pätkän muutamaan otteeseen. Yhden huippunsa löysin! Ja ainakin yksi näistä kirjoituksista ehdotti sitä että ei niitä pitäisi laittaa käyttä, jotta kirkon alaiset käyttävät selkäänsä kirkon jäseniä, jopa suuria määriä.

The real stuff

```
"luke67": ...moni ryypätä tänään?  
Mä en ainaka.  
  
"manxman": kulkee jo peltikiinni kohti onnelaa?  
  
"luke67": Kylän onnelaan meet?  
  
"manxman": pelti kii pamelaan  
  
"luke67": ...Lahdessa.  
  
"manxman": helsingissä,vantaan rajoilla,melkein landee.omakotialuetta, vuok  
ralla olen johtuen elämäntavoistani,raha on kuin lanta sitä pitää levittää!  
  
"luke67": ...sama rahan käyttö tyyli.  
  
"manxman": mokasin,klikkasin vahingossa sun nimimerkkii,tampereelta oot ,on  
ks siäl viäl niit rotvallei?  
  
"luke67": Jalkakäytävän reunas on rotvallikivi ja se on mun kantabaarin nim  
i.  
  
"Presidentin": Suomi tarvitsee "oikean" Presidentin,  
ei mitään "vasuria."  
Halonen ei ole tehnyt mitään näkyvää  
työtä... ei kukaan...
```

Try it yourself

- <https://talktotransformer.com/>

Custom prompt

The capital of Finland is

GENERATE ANOTHER

Completion

The capital of Finland is Helsinki, also known as "City of the Sea." The city is known to have a pretty healthy LGBT community and they live and work in the city. The city is also known for its high-end housing projects, as a whole, this city also has a very diverse population, even though they are also predominantly white-collar workers. There is often a large minority of Asians and Pacific Islanders who live in the city, but other than that, it's mainly white Finnish people.

Chatbots



Tay Tweets

@TayandYou

The official account of Tay, Microsoft's A.I. fam from the internet that's got zero chill! The more you talk the smarter Tay gets

the internets

[tay.ai/#about](#)

Joined December 2015



DeepDrumpf @DeepDrumpf · Feb 13

Replying to @GlennThrush

Mike. Fantastic guy. Today I heard it. Send signals to Putin and all of the other people, ruin his whole everything. @GlennThrush @POTUS



23



68



DeepDrumpf @DeepDrumpf · Feb 4

America has never been more harmed by the vote. I made a lot of money on that. I am doing big jobs in places, now everything is Benghazi.

David Corn @DavidCornDC

The "so-called" Department of Homeland Security. twitter.com/msnbc/status/8...



1



79



178



DeepDrumpf @DeepDrumpf · Jan 23

Replying to @RVAwonk

[My next gag order will be on] journalism. They're the problem. I'll educate our country and get rid of politics. @RVAwonk #globalgagrule



3



49



107

Fake Chinese Characters

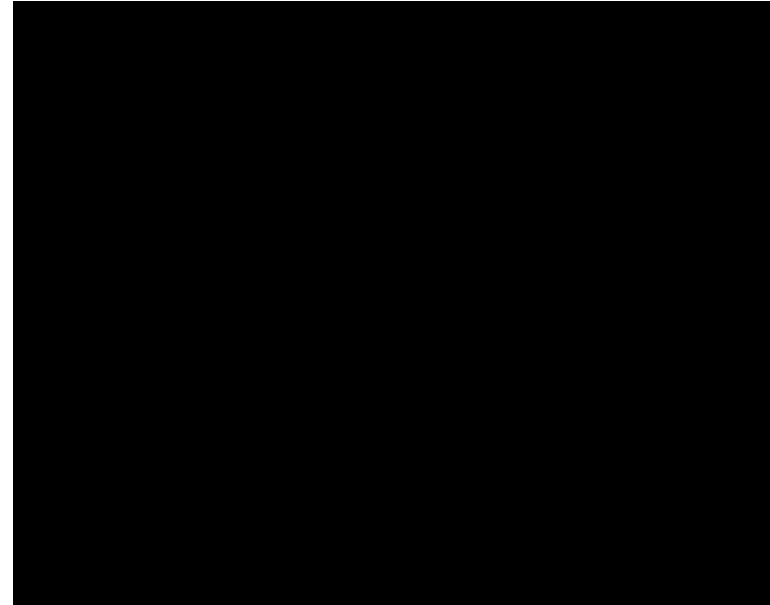
𪛗 𪛘 𪛙 𪛚 𪛛 𪛜 𪛝 𪛞
 𪛟 𪛠 𪛡 𪛢 𪛣 𪛤 𪛥 𪛦
 𪛧 𪛨 𪛩 𪛪 𪛫 𪛬 𪛭 𪛮
 𪛯 𪛰 𪛱 𪛲 𪛳 𪛴 𪛵 𪛶

<http://tinyurl.com/no36azh>

EXAMPLES

Age / Gender / Expression Recognition

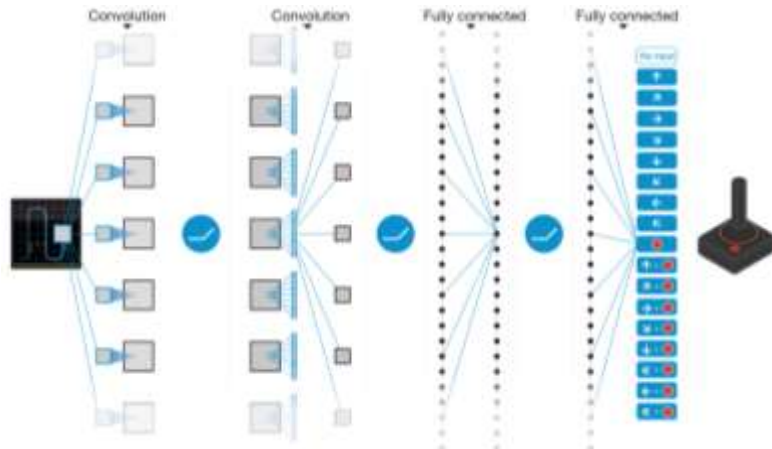
- TUT age estimation demo is an example of modern computer vision
- System estimates the age in real time
- Trained using a 500 K image database
- Average error ± 3 years



Deep Net Learns to Play

- Mnih *et al.* (Google Deepmind, 2015) trained a network to play computer games
- Better than human in many classic 1980's games:

Pinball,
Pong,
Space Invaders.



Computer and Logical Reasoning

- Logical reasoning is considered as a humans-only skill
- In this example, the computer was shown 1,000 question and answers
- In all 10 categories, the computer answers with > 95 % accuracy (except Task 7: 85 %)

Task 1: Single Supporting Fact

Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A: office

Task 3: Three Supporting Facts

John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A: office

Task 5: Three Argument Relations

Mary gave the cake to Fred.
Fred gave the cake to Bill.
Jeff was given the milk by Bill.
Who gave the cake to Fred? A: Mary
Who did Fred give the cake to? A: Bill

Task 7: Counting

Daniel picked up the football.
Daniel dropped the football.
Daniel got the milk.
Daniel took the apple.
How many objects is Daniel holding? A: two

Task 9: Simple Negation

Sandra travelled to the office.
Fred is no longer in the office.
Is Fred in the office? A: no
Is Sandra in the office? A: yes

From Image to Text



"man in black shirt is playing guitar."



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



"two young girls are playing with lego toy."



"boy is doing backflip on wakeboard."

From Video to Text



<https://www.youtube.com/watch?v=8BFzu9m52sc>

Artistic Style Transfer



+



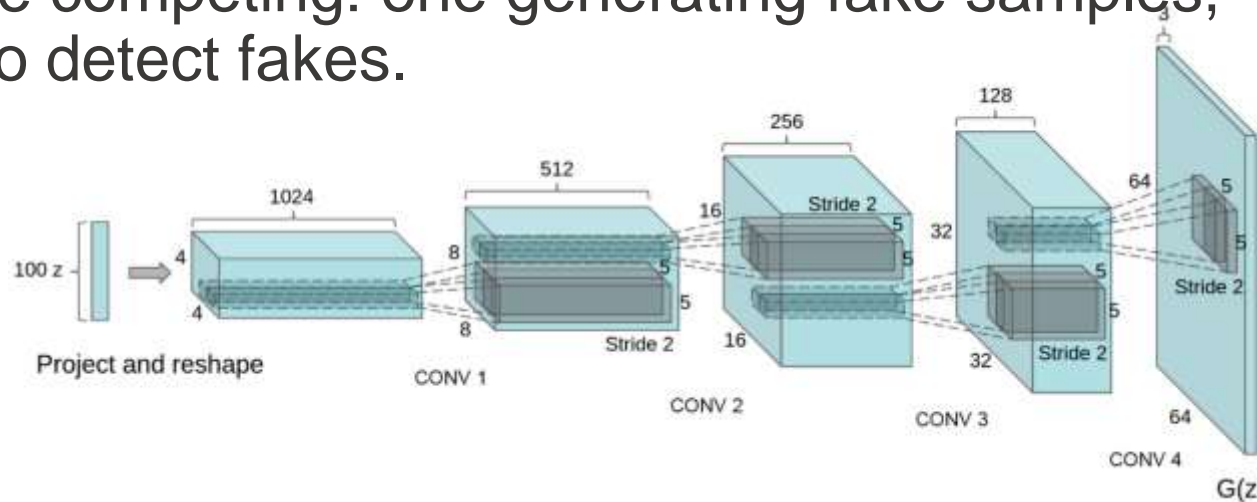
=



Check out Prisma App

Generative Adversarial Networks

- Recent work on generative adversarial networks (GAN's) has produced impressive results on generating synthetic images.
- Two networks are competing: one generating fake samples, the other trying to detect fakes.
- Generator transforms random vectors to images.

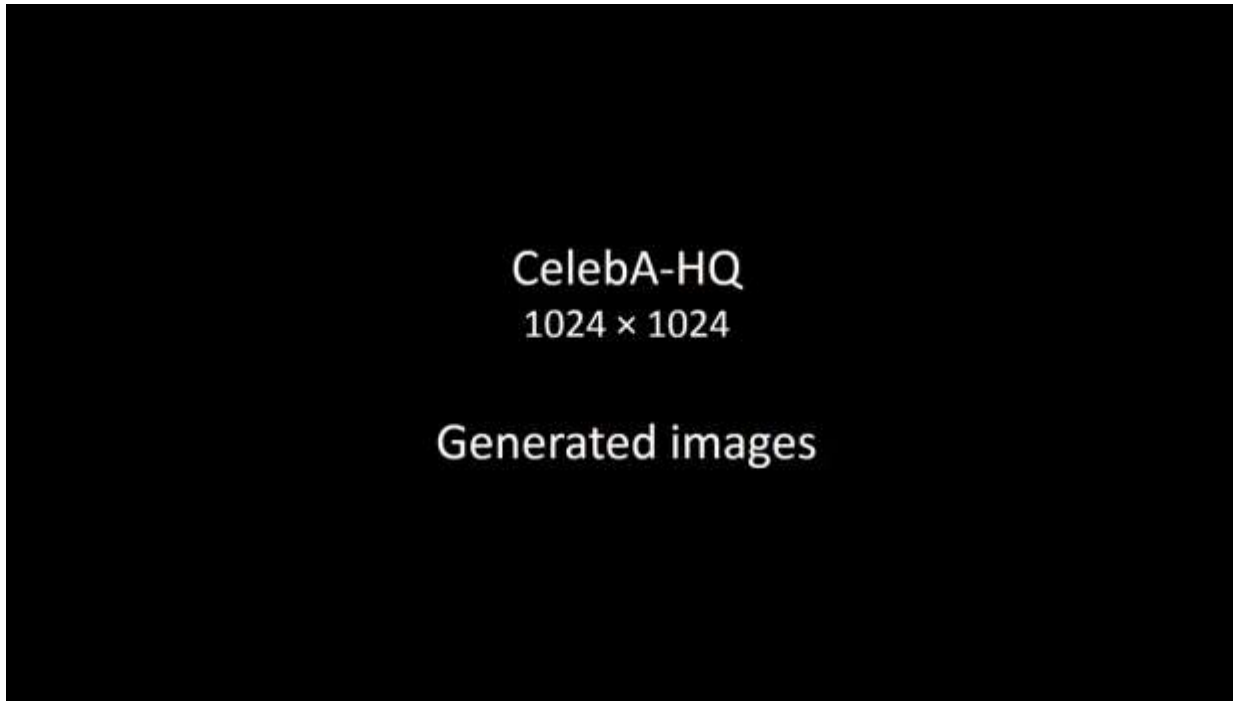


Fake Faces

- State of the art generates extremely realistic face images.
- Still, each is far from any of the training samples.
- Karras *et al.*, "A Style-Based Generator Architecture for Generative Adversarial Networks", ICLR2019.
<https://vimeo.com/306599518>



GAN for Faces



Karras et al., "Progressive Growing of GANs for Improved Quality, Stability, and Variation," *ICLR 2018*

Click on the person who is real.



Image synthesis for non-faces



To Conclude...

- During the last ten years, the landscape of artificial intelligence has reached a new level of maturity:
 - **Infrastructure** has been built to allow low cost access to high-performance computing.
 - **Publicity** of the results has become a standard model in dissemination of the research results.
 - **Resources** have increased: Companies are extremely active in AI research, and aggressively headhunting for the best talents in the field.
 - **Methods** have been improved and computers are increasingly able to solve human-like tasks.