



Intro to Deep Learning

Computer Vision

Carnegie Mellon University (Kris Kitani)

Recall:

‘Classical’

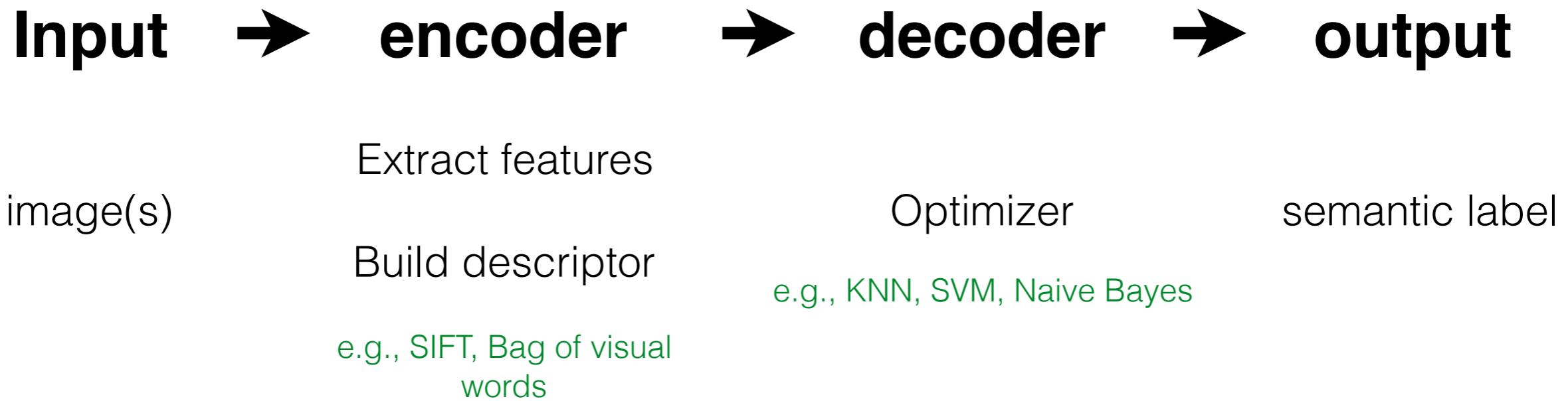
Image Classification Pipeline



Recall:

‘Classical’

Image Classification Pipeline



Recall:

‘Classical’

Image Classification Pipeline

Input → **encoder** → **decoder** → **output**

```
graph LR; A[image(s)] --> B[Extract features]; B --> C[Build descriptor<br/>e.g., SIFT, Bag of visual words]; C --> D[Optimizer<br/>e.g., KNN, SVM, Naive Bayes]; D --> E[semantic label]
```



Learned from the data

Recall:

‘Classical’

Image Classification Pipeline

Input → **encoder** → **decoder** → **output**

```
graph LR; A[image(s)] --> B[Extract features]; B --> C[Optimizer]; C --> D[semantic label]
```

Build descriptor

e.g., SIFT, Bag of visual words



Designed by you

Painful.

Optimizer

e.g., KNN, SVM, Naive Bayes

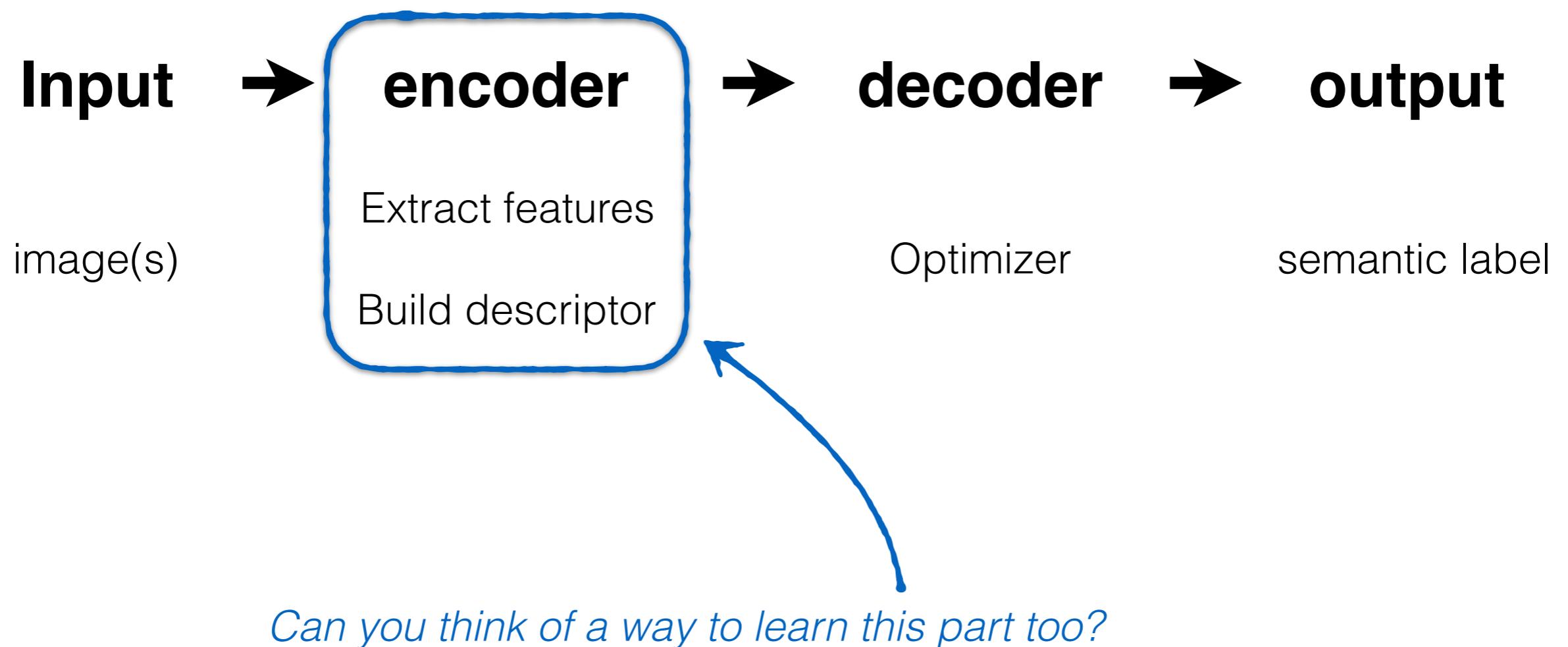


Learned from the data

Recall:

‘Classical’

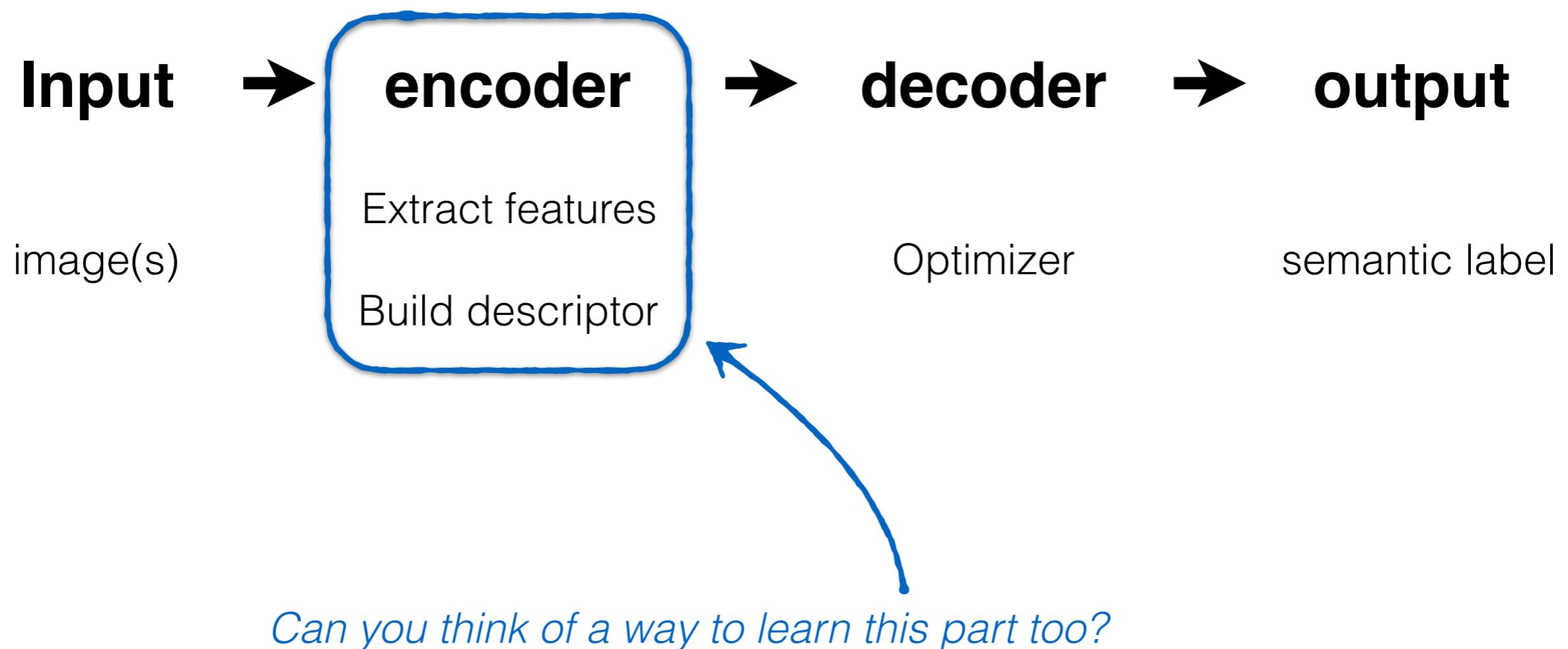
Image Classification Pipeline



Recall:

‘Classical’

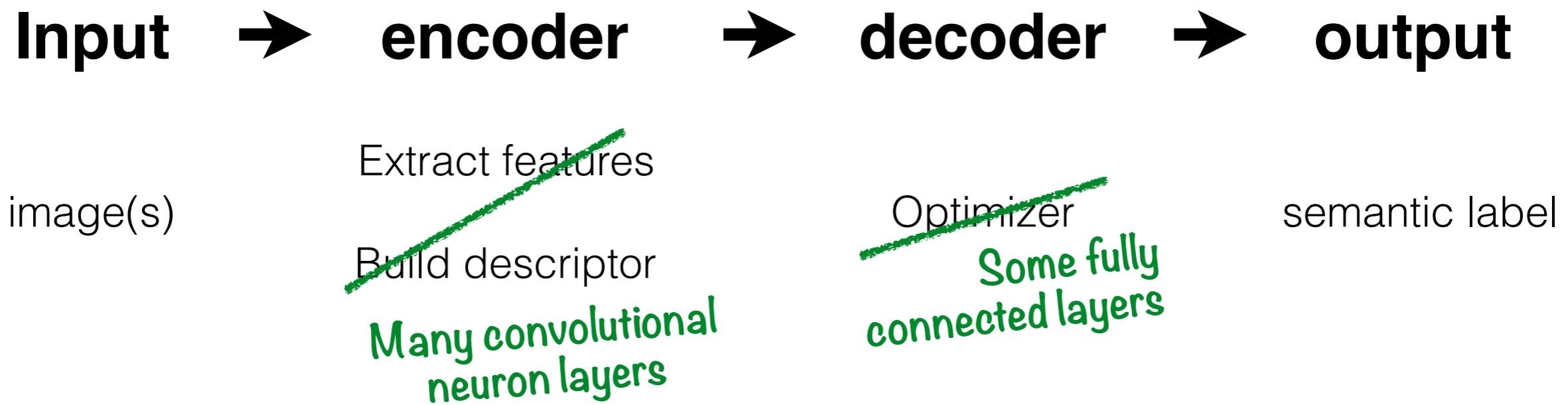
Image Classification Pipeline



*How can we learn it all ‘**end to end**?’*

'modern'
~~**'Classical'**~~ (revival of a classical)

Image Classification Pipeline

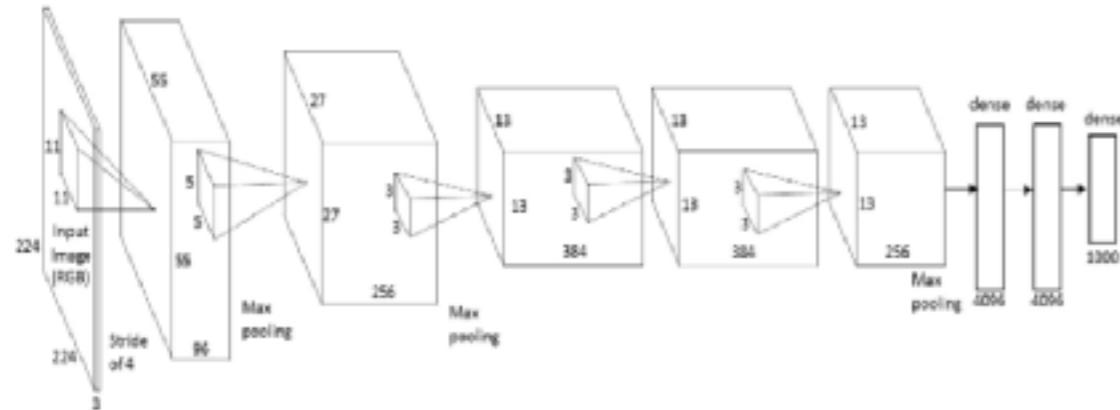


'Deep Network'

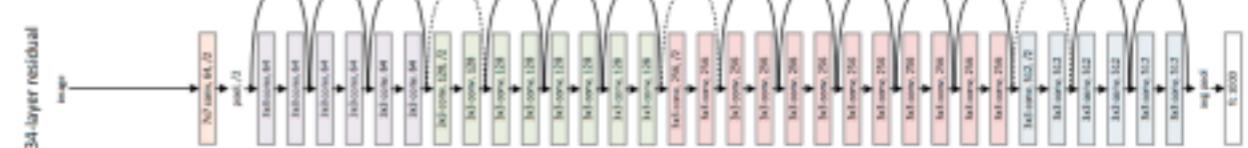
Use gradient descent to learn everything
(except for the architecture, which you have to design. painful.)

Popular CNN architectures

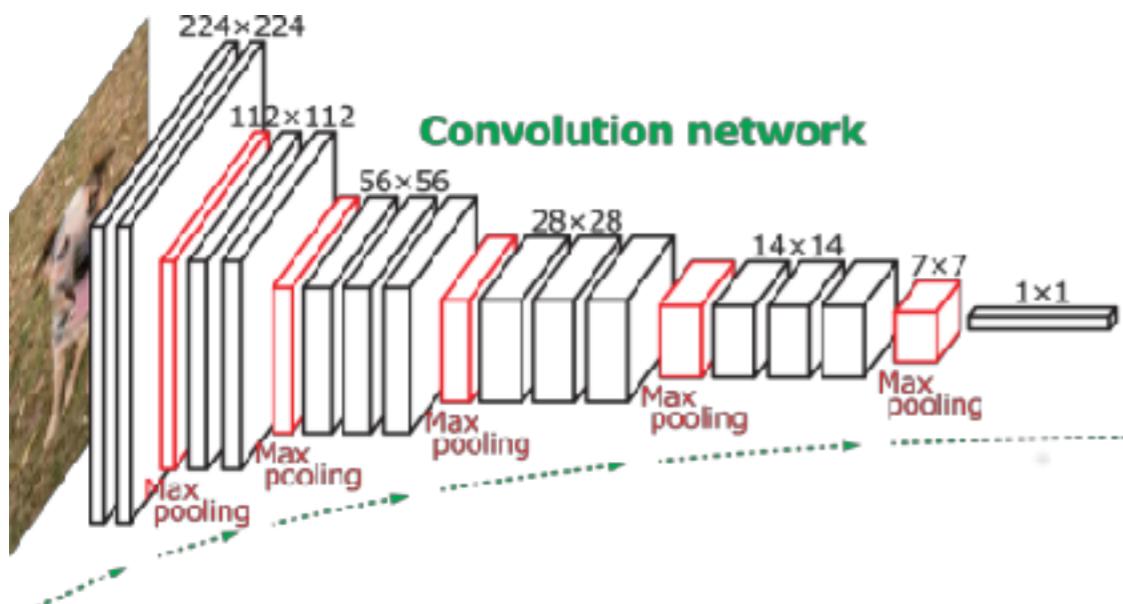
Covered in this session



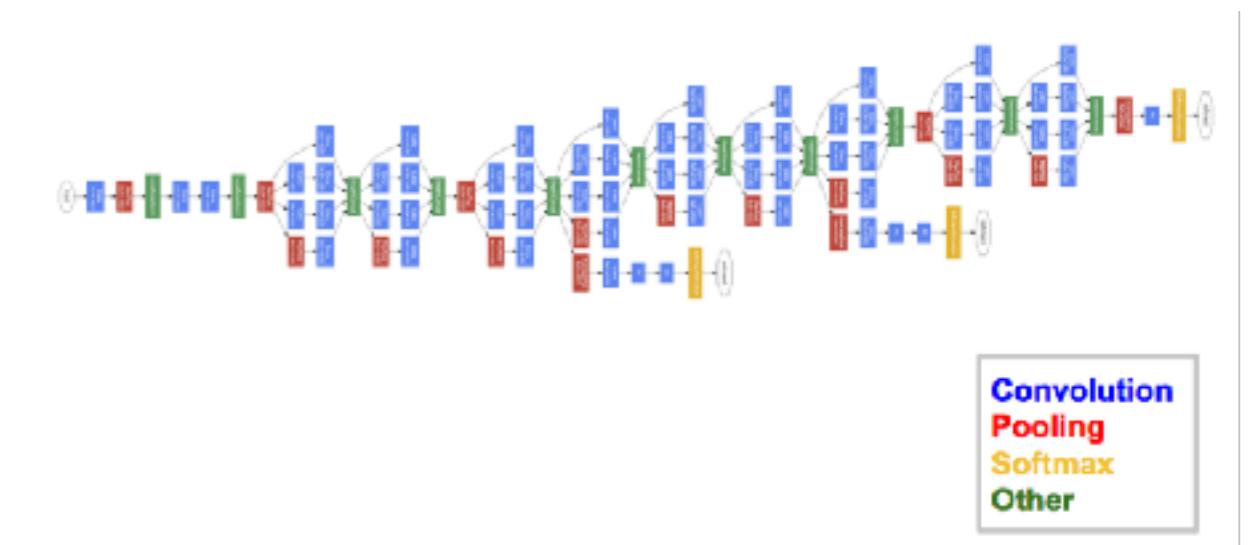
AlexNet



ResNet



VGG



GoogLeNet

But first some recent results of
deep learning in computer vision ...

DeepFace: Closing the Gap to Human-Level Performance in Face Verification

Yaniv Taigman

Ming Yang

Marc'Aurelio Ranzato

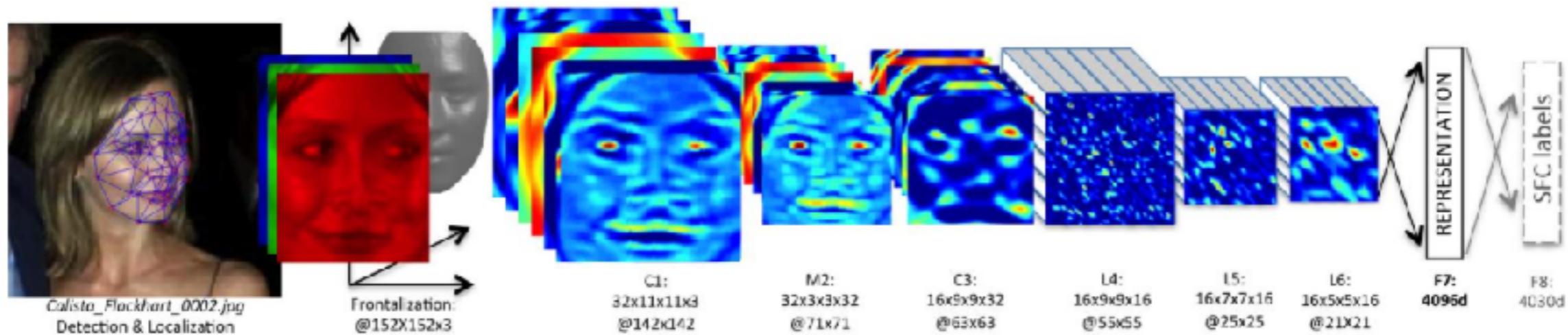
Lior Wolf

Facebook AI Research
Menlo Park, CA, USA

{yaniv, mingyang, ranzato}@fb.com

Tel Aviv University
Tel Aviv, Israel

wolf@cs.tau.ac.il



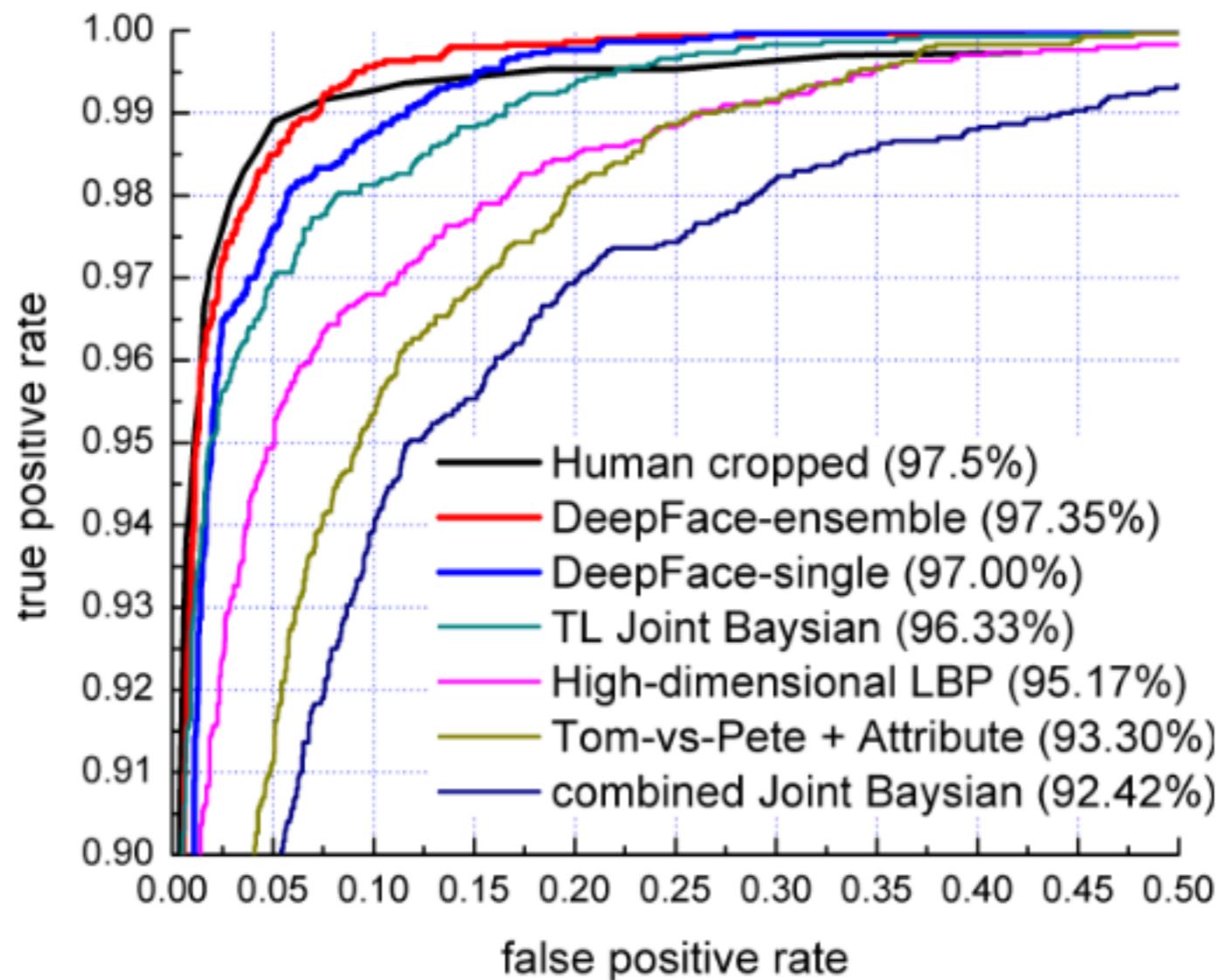


Figure 3. The ROC curves on the *LFW* dataset. Best viewed in color.

CNNs match human accuracy on face verification!



Google Research Blog

The latest news from Research at Google

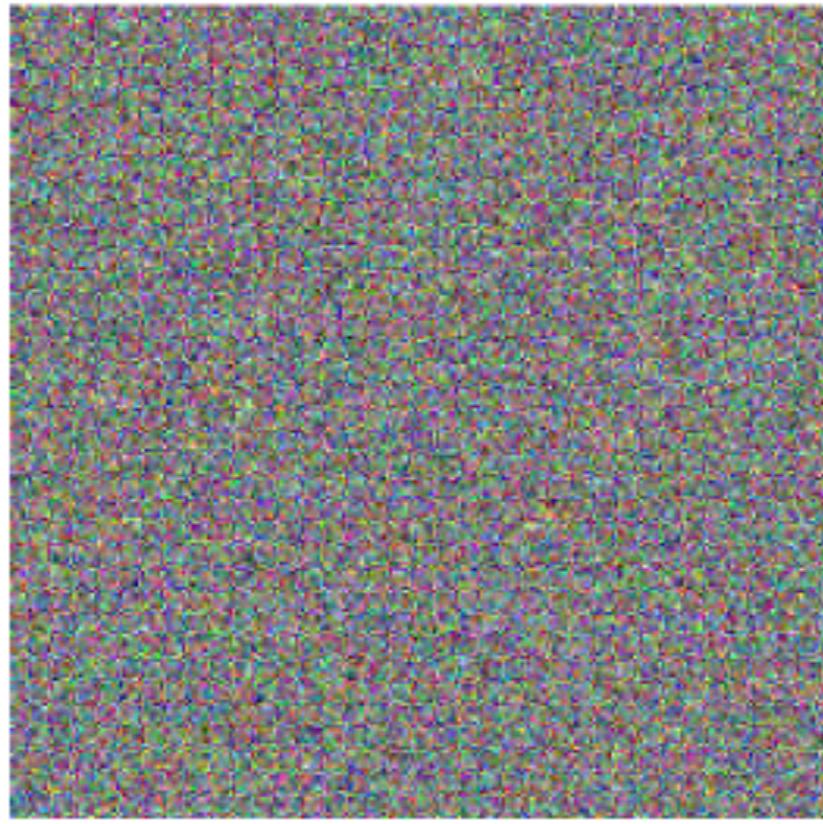
Inceptionism: Going Deeper into Neural Networks

Wednesday, June 17, 2015

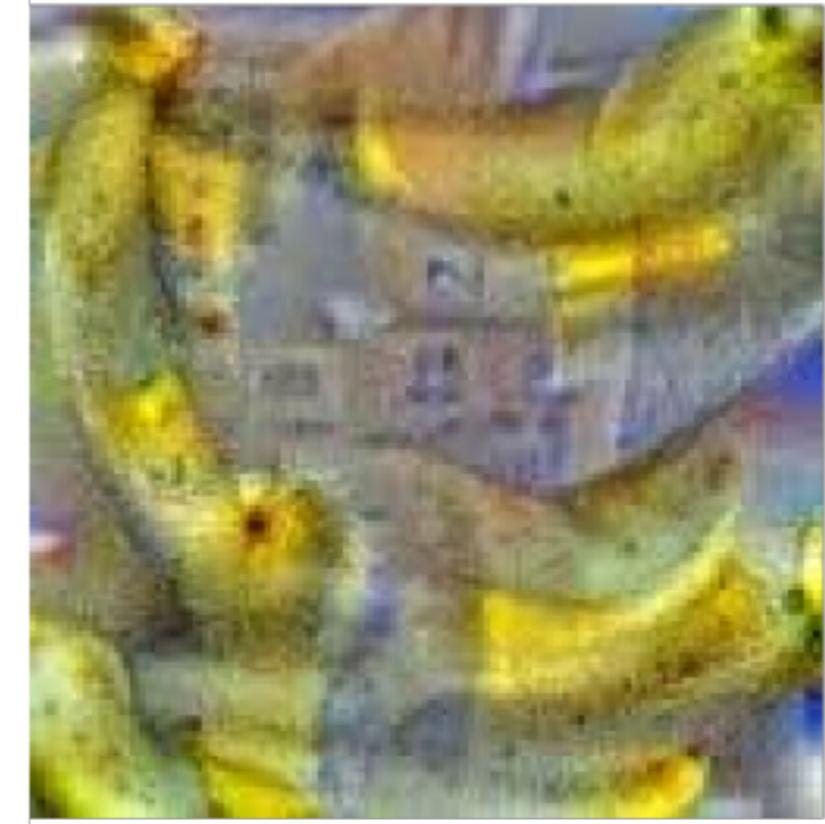
Posted by Alexander Mordvintsev, Software Engineer, Christopher Olah, Software Engineering Intern and Mike Tyka, Software Engineer

Update - 13/07/2015

Images in this blog post are licensed by Google Inc. under a [Creative Commons Attribution 4.0 International License](#). However, images based on places by [MIT Computer Science and AI Laboratory](#) require additional permissions from MIT for use.



optimize
with prior



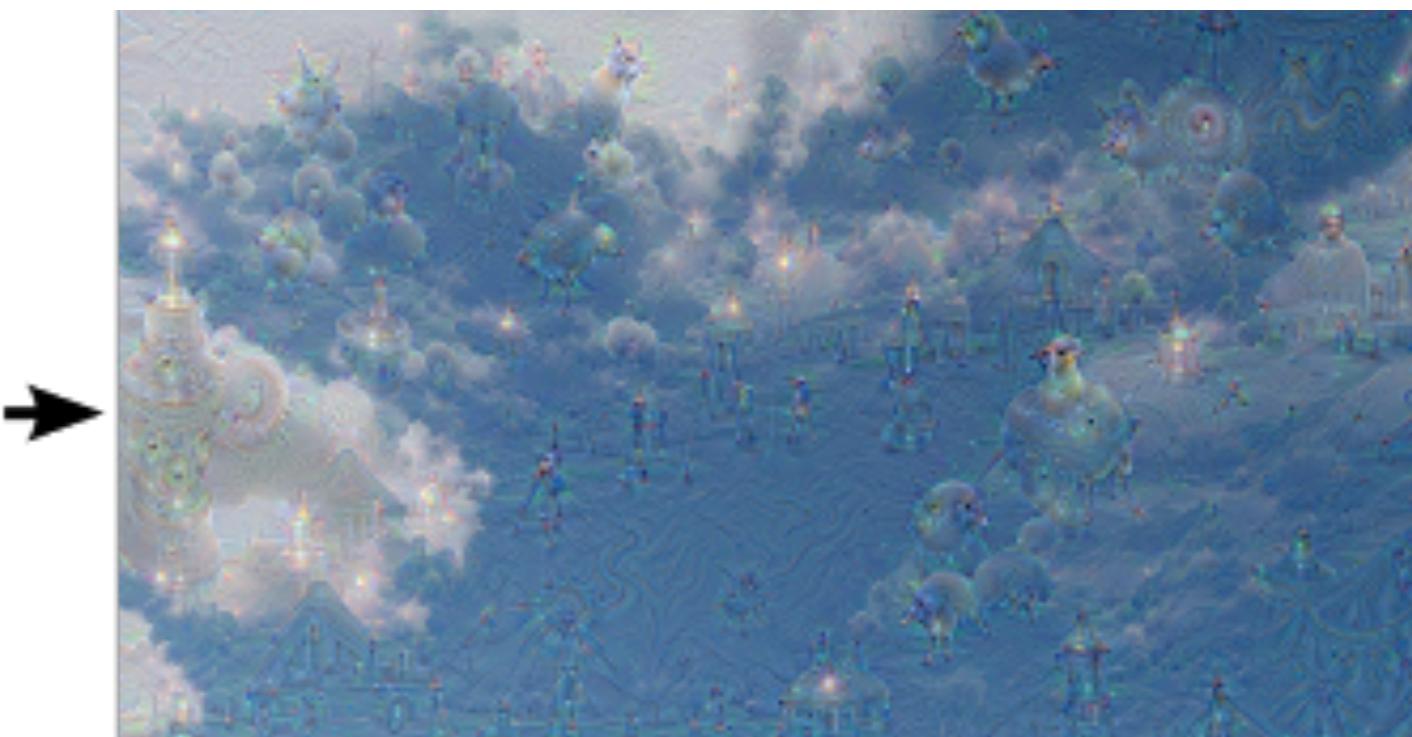
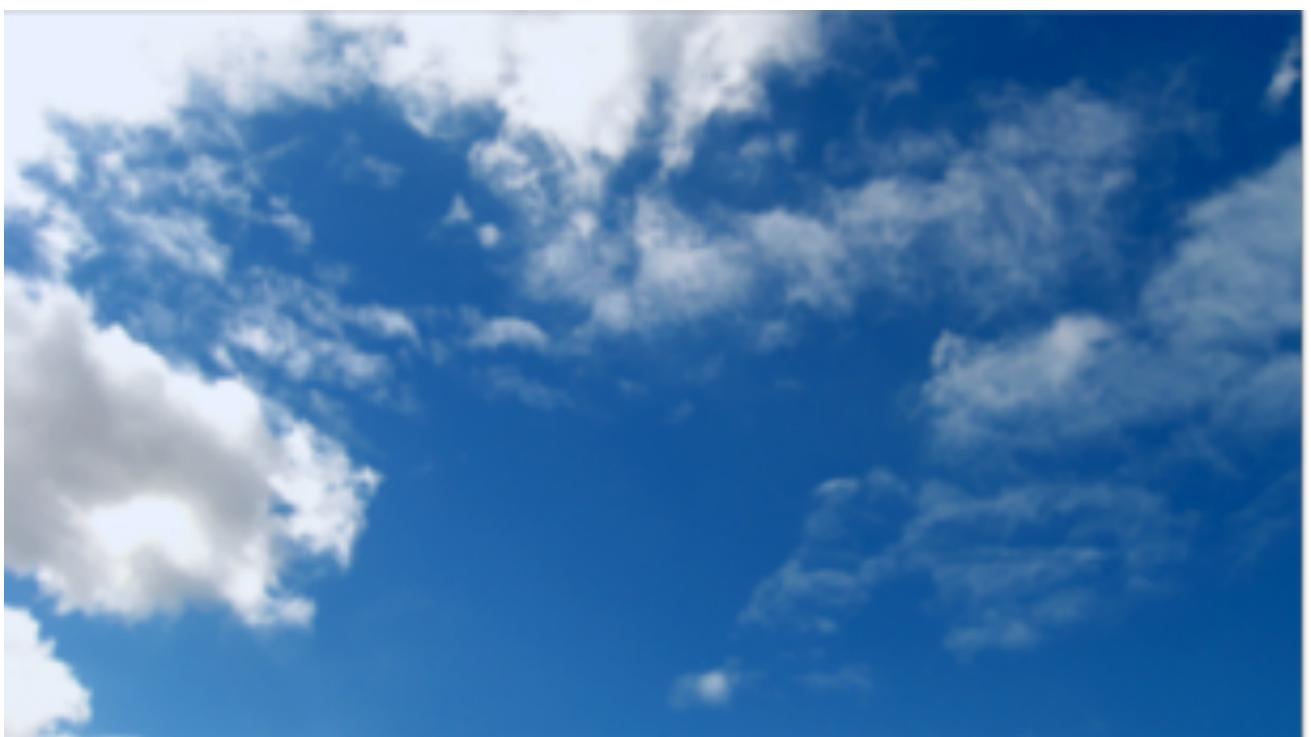
gradient updates only to $\frac{\partial \mathcal{L}}{\partial \mathbf{x}}$ to minimize ‘banana’ loss

Deep Dream Generator

Generate your own deep dream photos and images for free.

 UPLOAD IMAGE

 LOG IN



"Admiral Dog!"



"The Pig-Snail"



"The Camel-Bird"



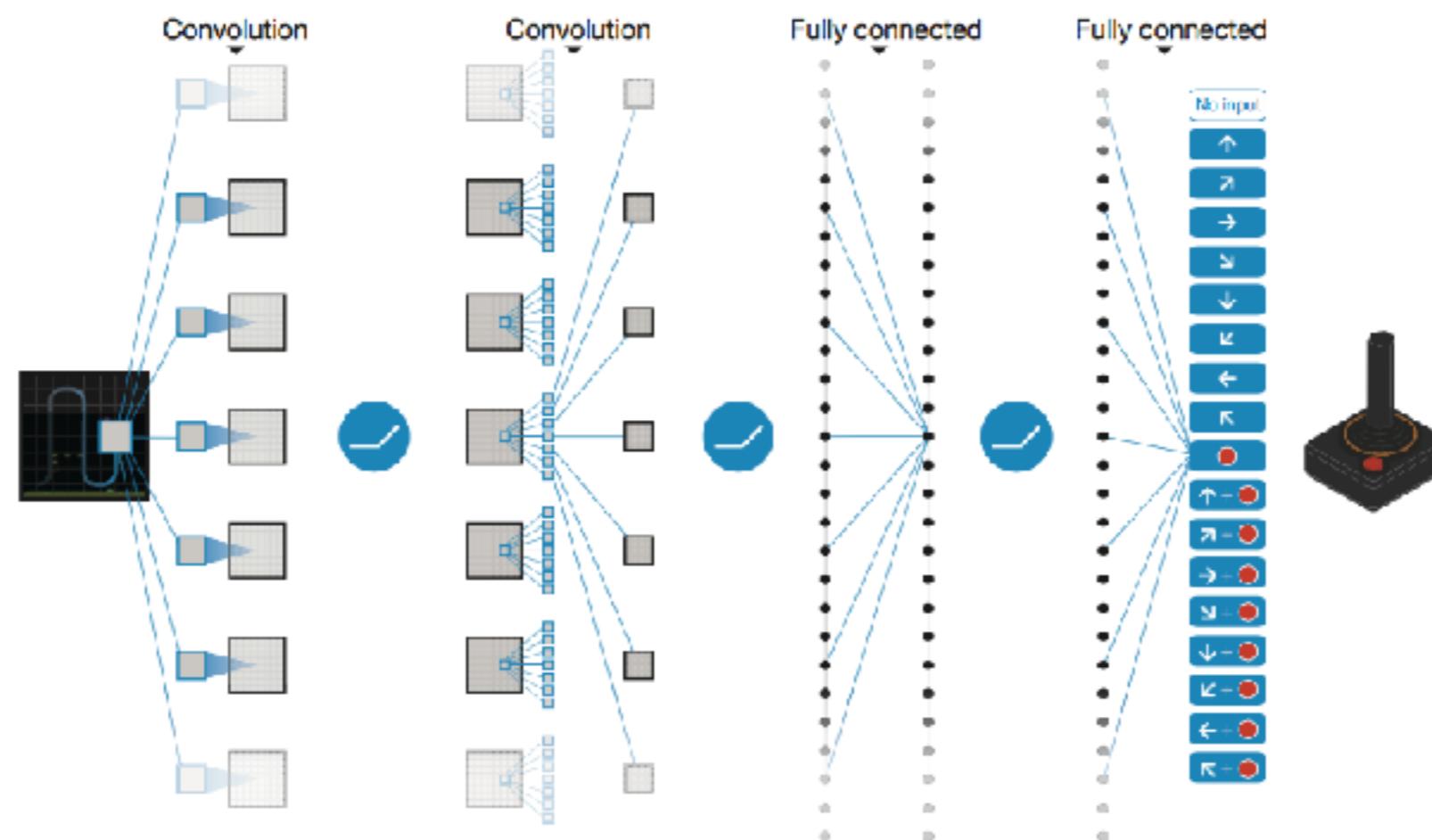
"The Dog-Fish"



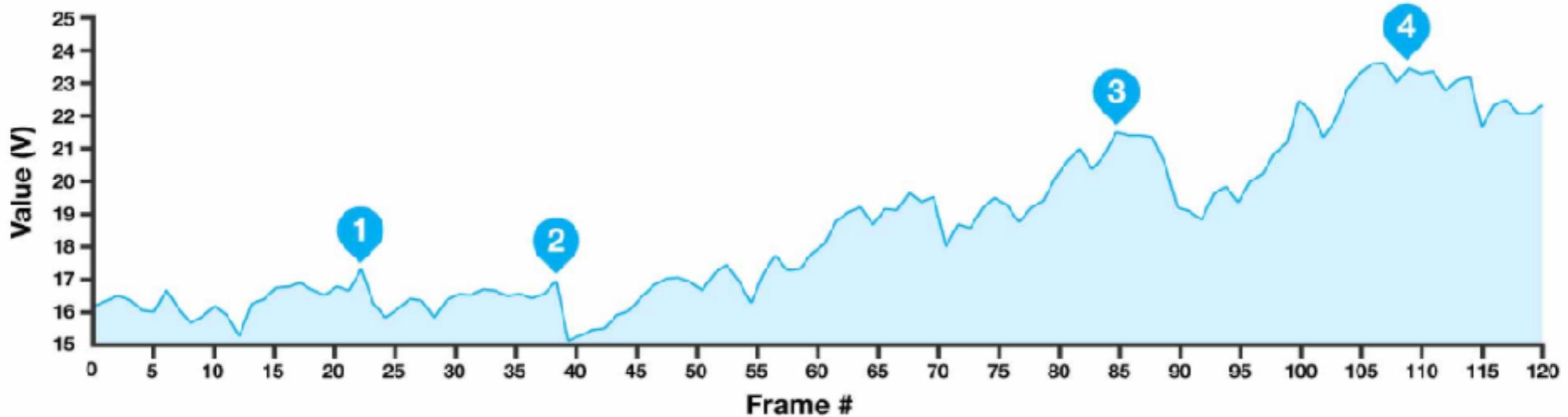
CNNs can create weird hallucinations!

Human-level control through deep reinforcement learning

Volodymyr Mnih^{1*}, Koray Kavukcuoglu^{1*}, David Silver^{1*}, Andrei A. Rusu¹, Joel Veness¹, Marc G. Bellemare¹, Alex Graves¹, Martin Riedmiller¹, Andreas K. Fidjeland¹, Georg Ostrovski¹, Stig Petersen¹, Charles Beattie¹, Amir Sadik¹, Ioannis Antonoglou¹, Helen King¹, Dharshan Kumaran¹, Daan Wierstra¹, Shane Legg¹ & Demis Hassabis¹



a



CNNs can play video games!

But wait, that's not all...

End to End Learning for Self-Driving Cars

Mariusz Bojarski
NVIDIA Corporation
Holmdel, NJ 07735

Davide Del Testa
NVIDIA Corporation
Holmdel, NJ 07735

Daniel Dworakowski
NVIDIA Corporation
Holmdel, NJ 07735

Bernhard Firner
NVIDIA Corporation
Holmdel, NJ 07735

Beat Flepp
NVIDIA Corporation
Holmdel, NJ 07735

Prasoon Goyal
NVIDIA Corporation
Holmdel, NJ 07735

Lawrence D. Jackel
NVIDIA Corporation
Holmdel, NJ 07735

Mathew Monfort
NVIDIA Corporation
Holmdel, NJ 07735

Urs Muller
NVIDIA Corporation
Holmdel, NJ 07735

Jiakai Zhang
NVIDIA Corporation
Holmdel, NJ 07735

Xin Zhang
NVIDIA Corporation
Holmdel, NJ 07735

Jake Zhao
NVIDIA Corporation
Holmdel, NJ 07735

Karol Zieba
NVIDIA Corporation
Holmdel, NJ 07735

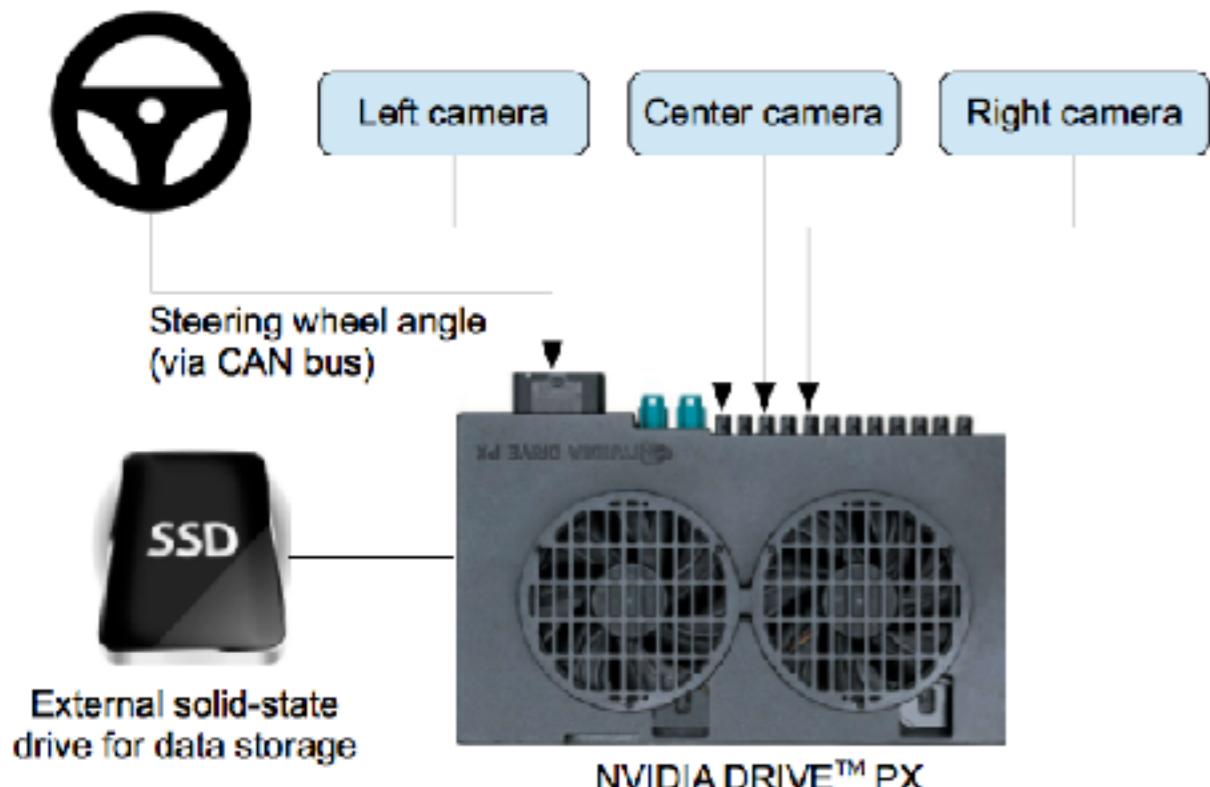


Figure 1: High-level view of the data collection system.

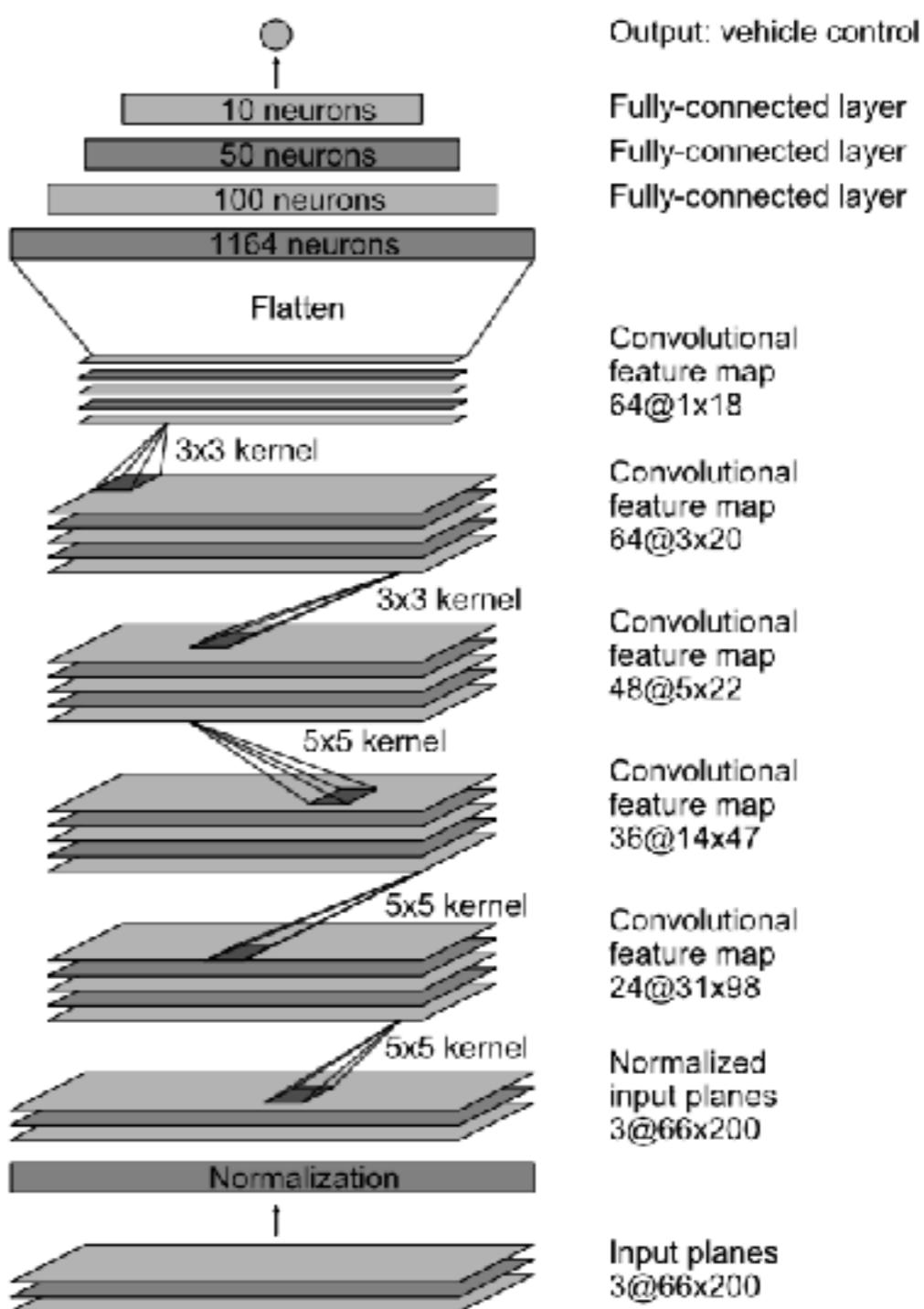


Figure 4: CNN architecture. The network has about 27 million connections and 250 thousand parameters.

DAVE 2 Driving a Lincoln

- A convolutional neural network
- Trained by human drivers
- Learns perception, path planning, and control
"pixel in, action out"
- Front-facing camera is the only sensor

Camera + CNNs can drive cars!

Released Jan 2018 from NVIDIA



<https://www.youtube.com/watch?v=Sm-NBdSzP6E>