计算机视觉班

The Last Lecture

Outline

- 课程总结
  - 深度学习的本质
  - 深度学习目前的进展
- 一个实例：鲸鱼识别

深度学习何处适用

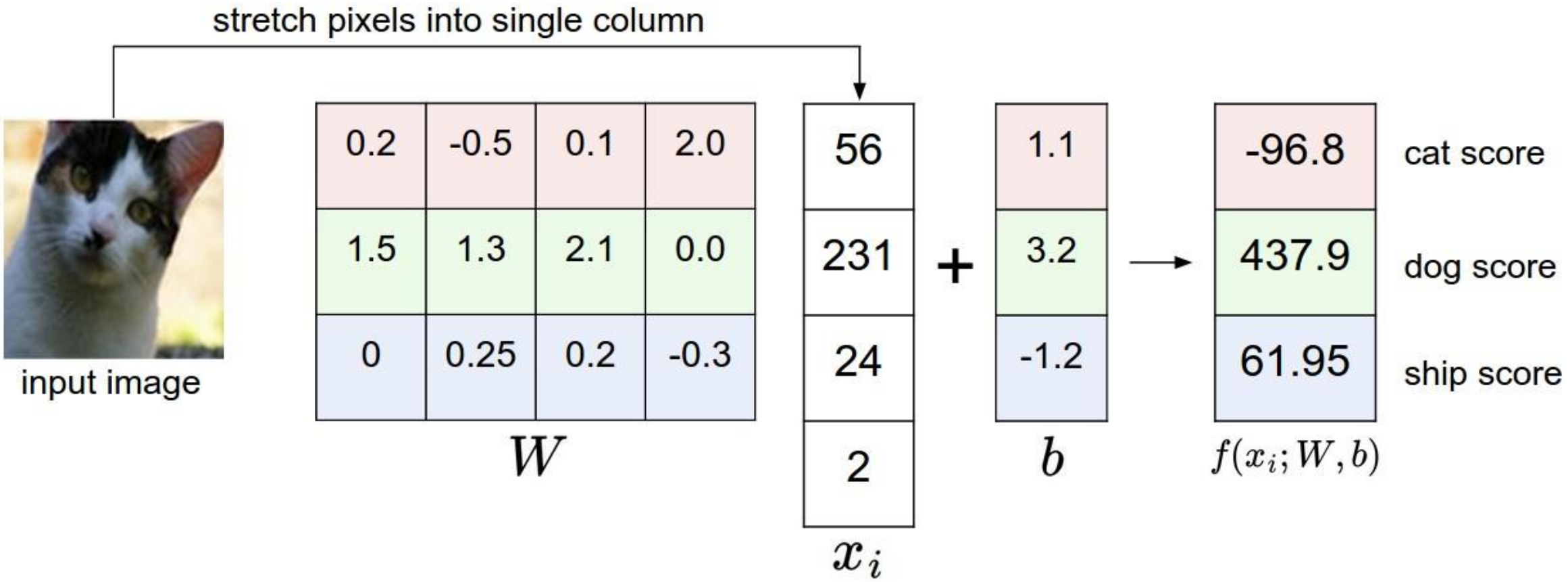
深度学习最重要的作用是表示学习

初始表示 与 合适表示 相距甚远时适用。

深度学习为何有效

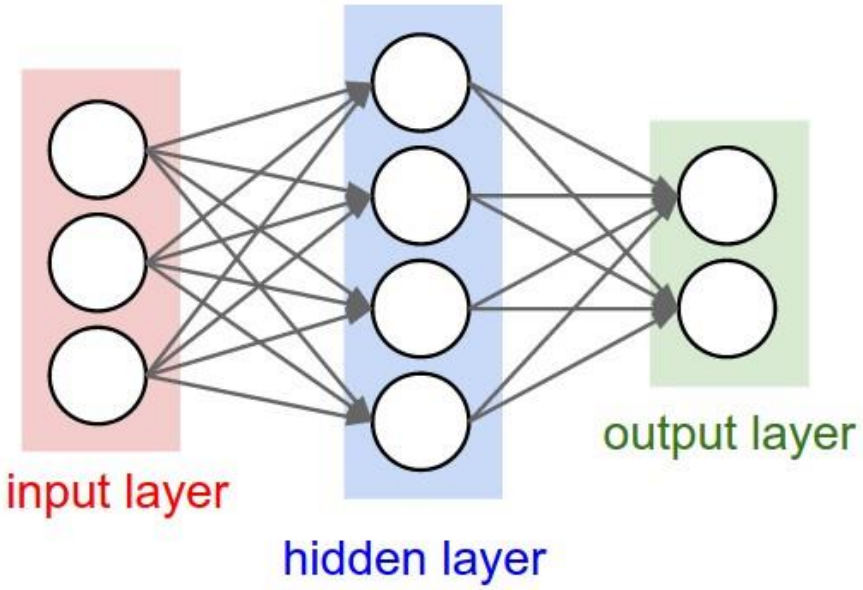
- 模型复杂度
- 大量数据
- 硬件崛起

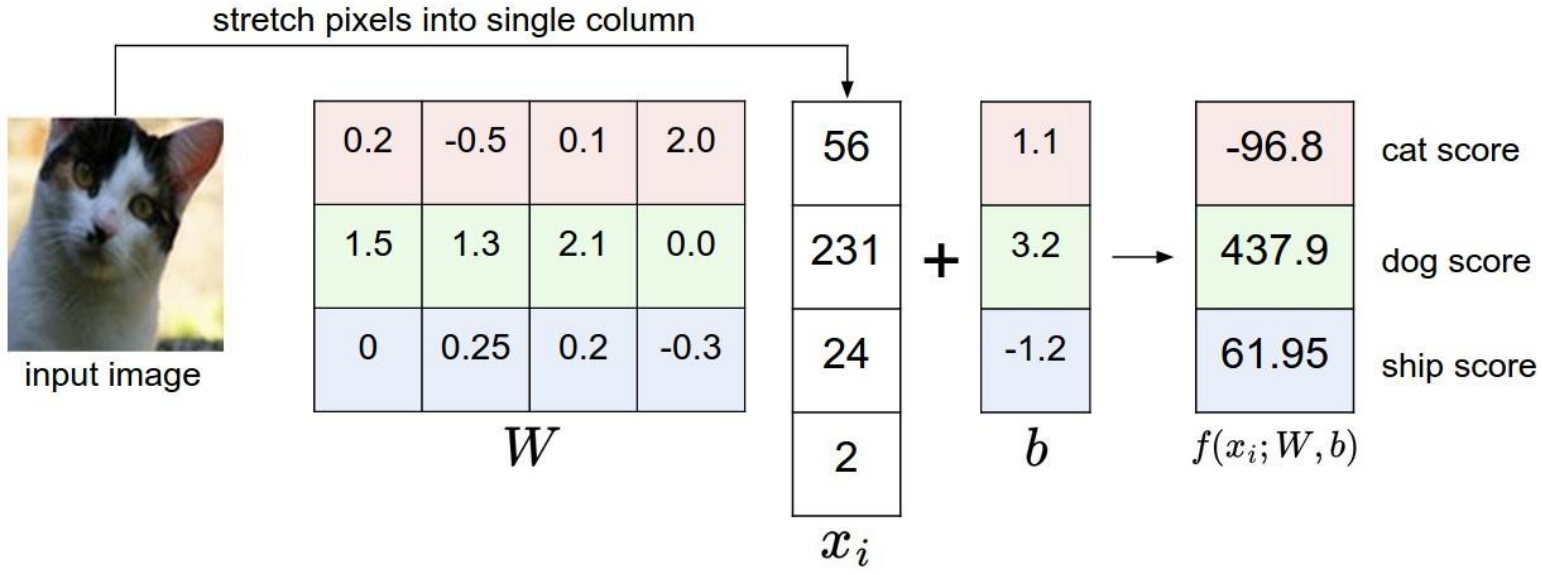
Linear Models



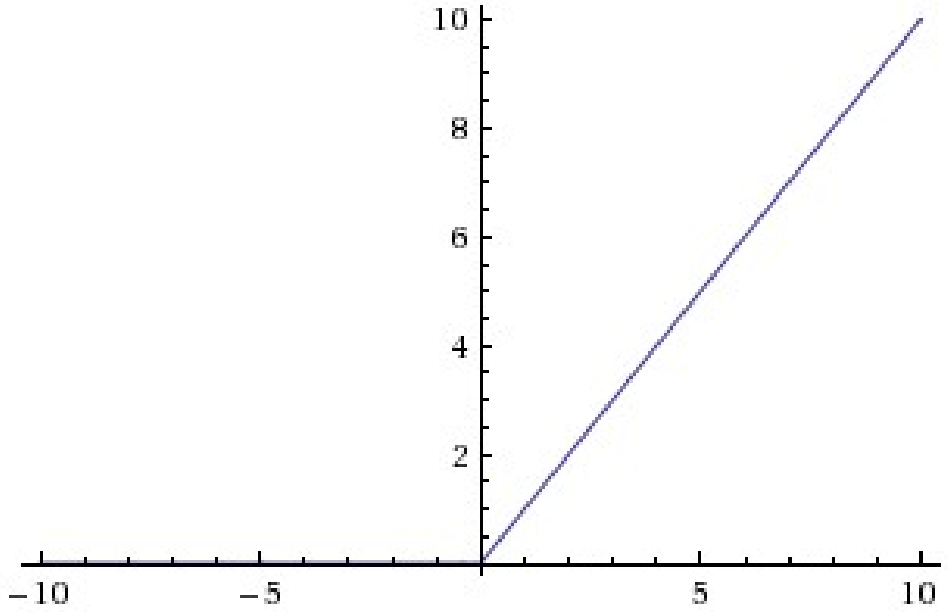
Linear Weights Visualized

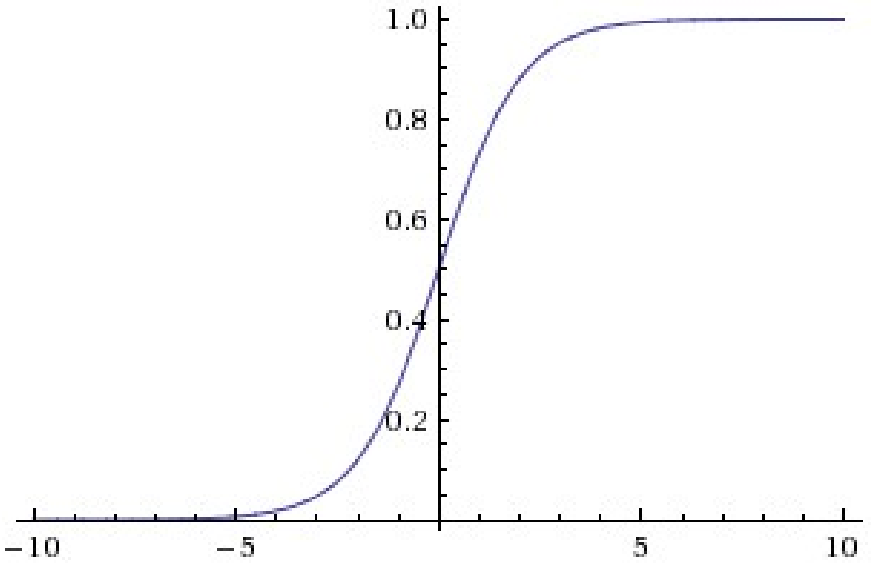
Now, Neural Networks

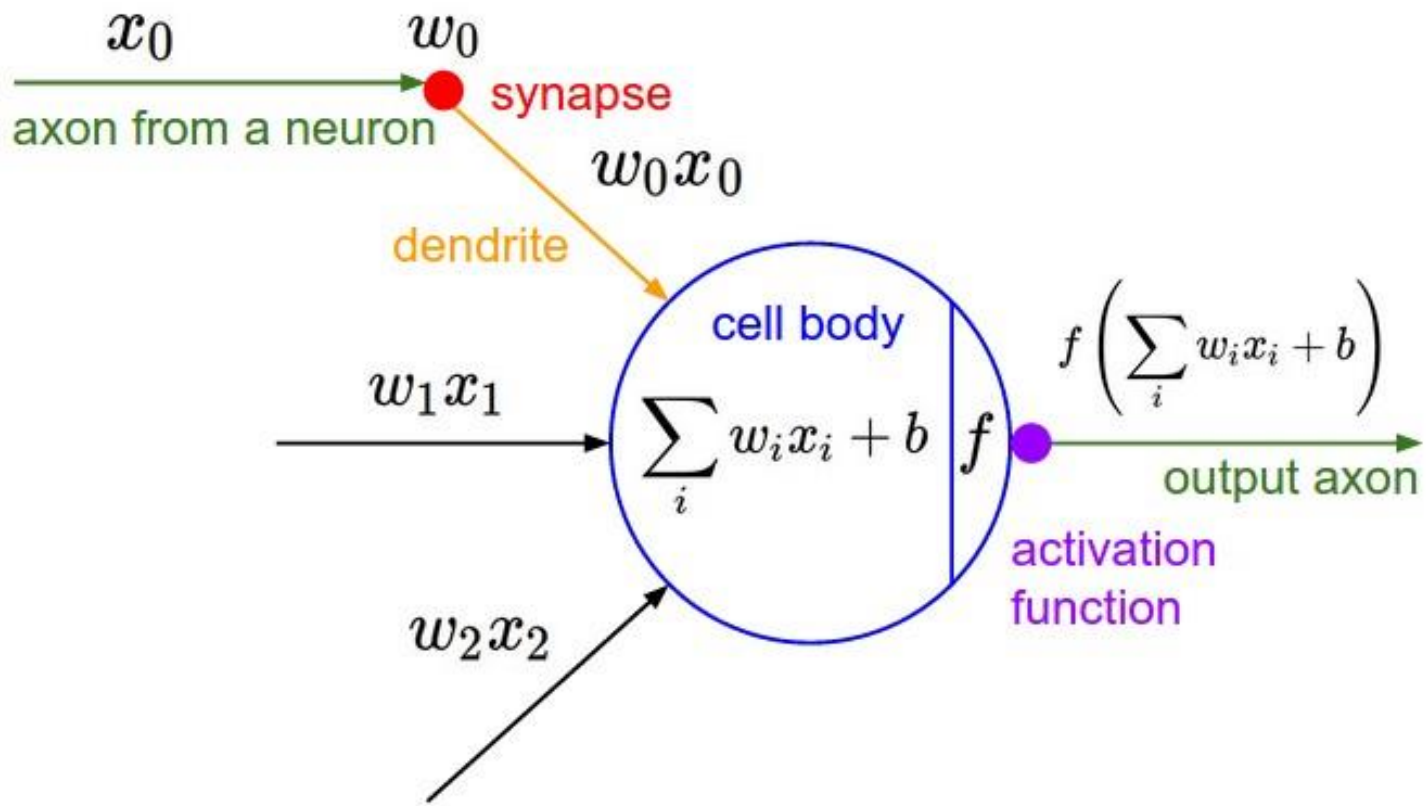




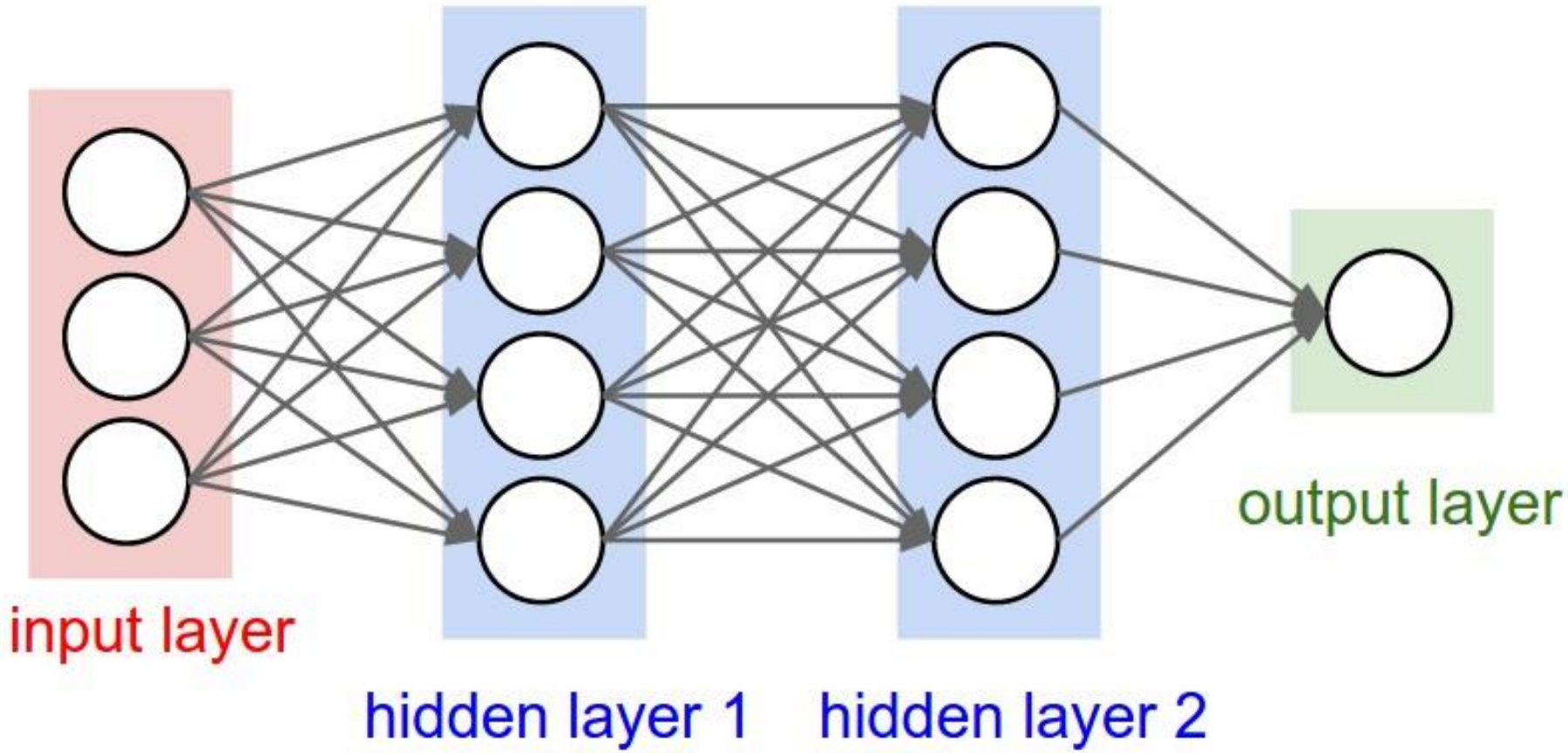
Inside a Neuron



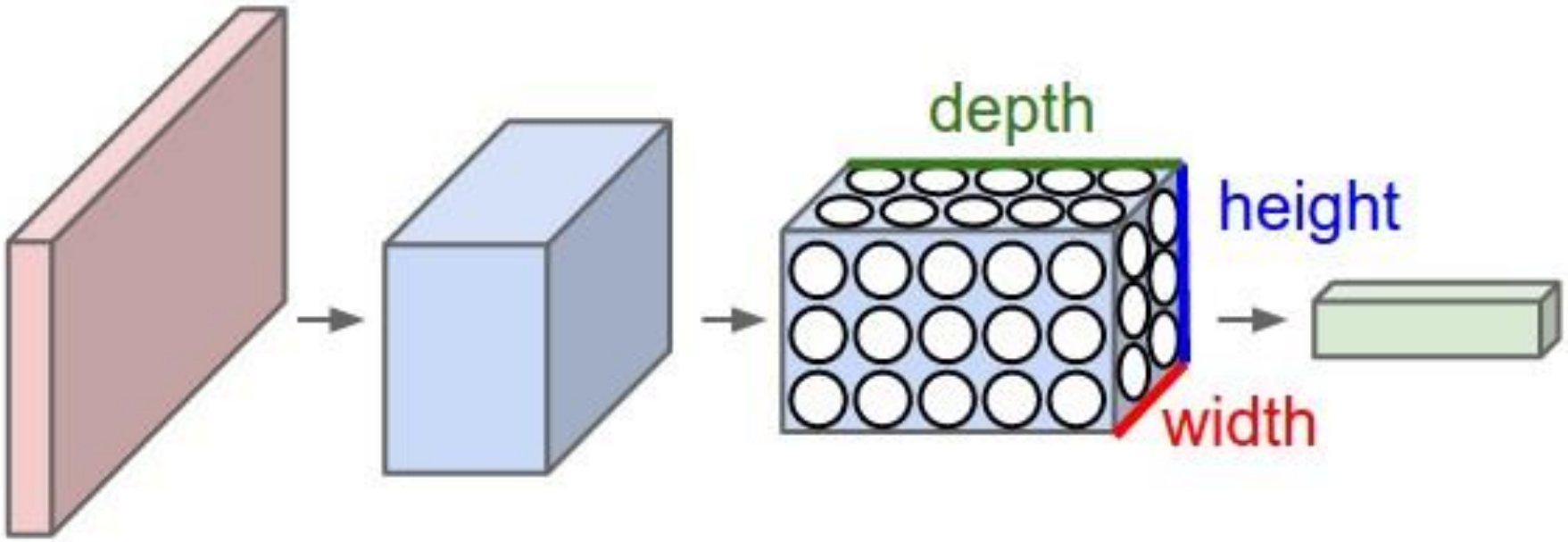



x_0
w_0
synapse
axon from a neuron
$w_0 x_0$
dendrite
cell body
$w_1 x_1$
$\sum_i w_i x_i + b$
f
$f\left(\sum_i w_i x_i + b\right)$
output axon
activation function
$w_2 x_2$

Multi-layer NN

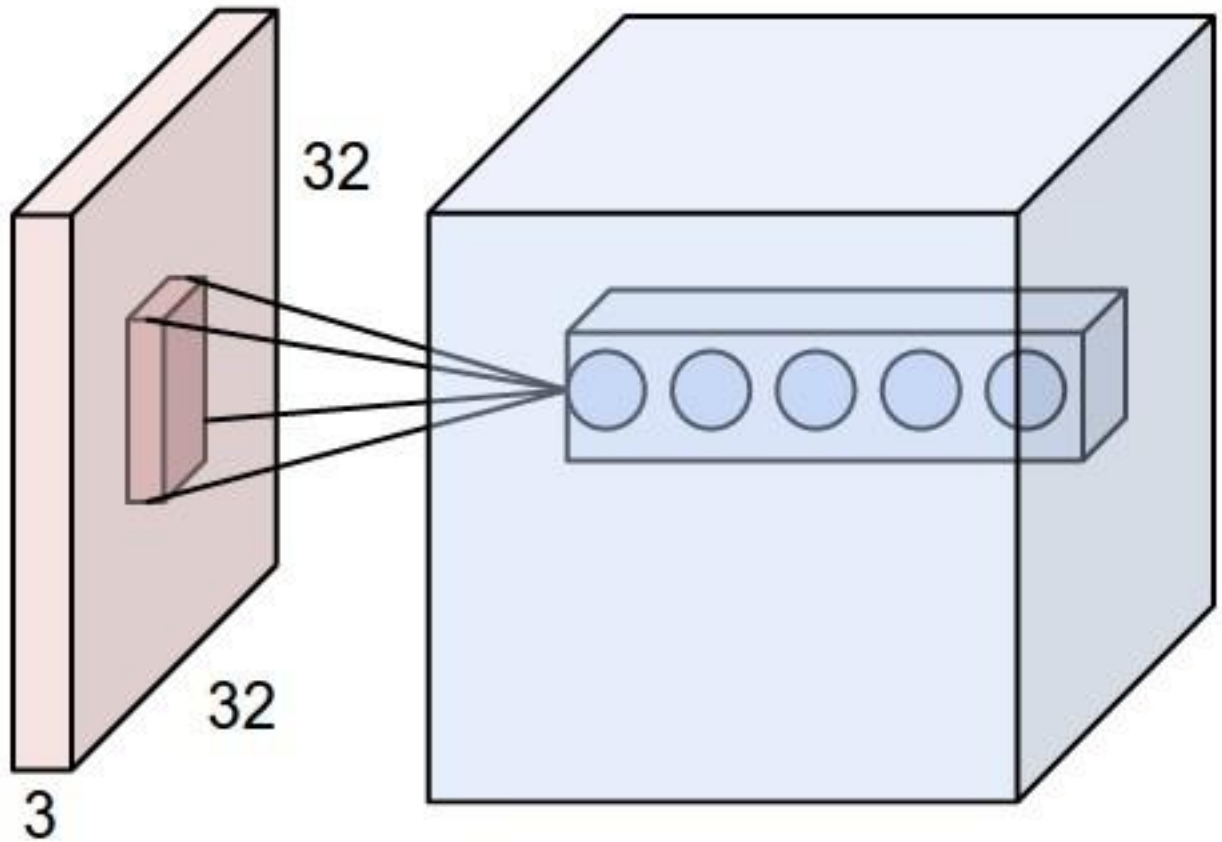


Convolutional Networks



Key ideas

1. Local connectivity
2. Shared Weights (parameter sharing)


32
32
3

神经网络发展回顾

- 40年代 萌芽期：Hebb 学习规则 1945
- 58-69 繁荣期 perceptron 1959 -------电子计算机
- 69年-80年 冰河期
- 85-95 繁荣期 -------x86系列/内存条
- 95-10 沉寂期
- 10-now 繁荣期 ------GPU通用计算

- 冷 10年 热15年

往哪儿走

- 鲁棒性：
  - 可重用
  - 可演进
  - 可了解
- Learnware = model + specification

计算机视觉和机器学习的关系

机器学习方法解决CV问题具有通用性
Eg：CNN/LSTM 不仅仅可以用来做CV

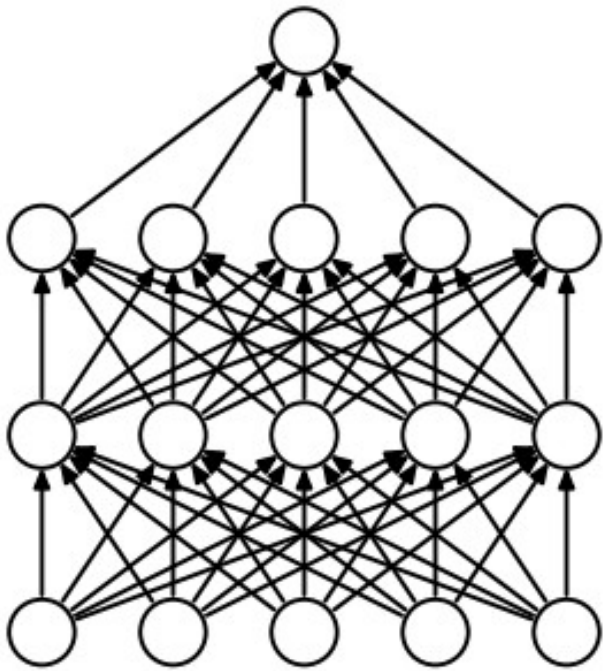
深度学习：（目前）最少领域知识的绝决方案

What makes AI great again

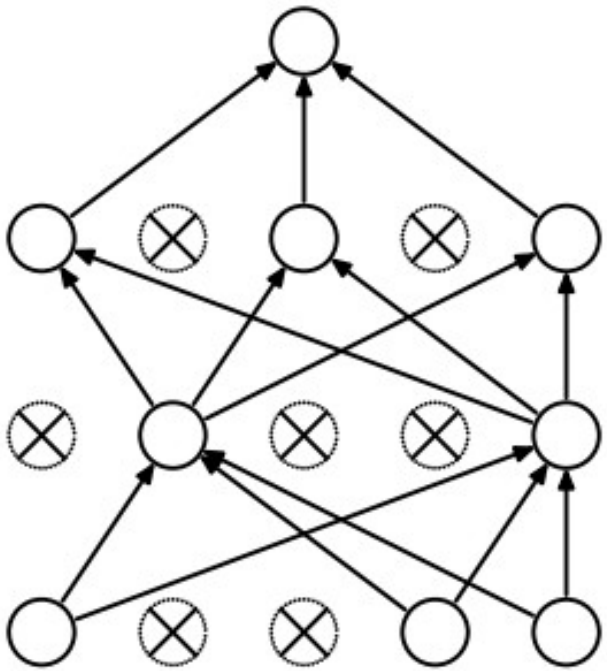
1. Compute (the obvious one: Moore's Law, GPUs, ASICs),
2. Data (in a nice form, not just out there somewhere on the internet - e.g. ImageNet),
3. Algorithms (research and ideas, e.g. backprop, CNN, LSTM), and
4. Infrastructure (software under you - Linux, TCP/IP, Git, TensorFlow,Theano, etc.).

The Dark Knowledge

Dropout



(a) Standard Neural Net

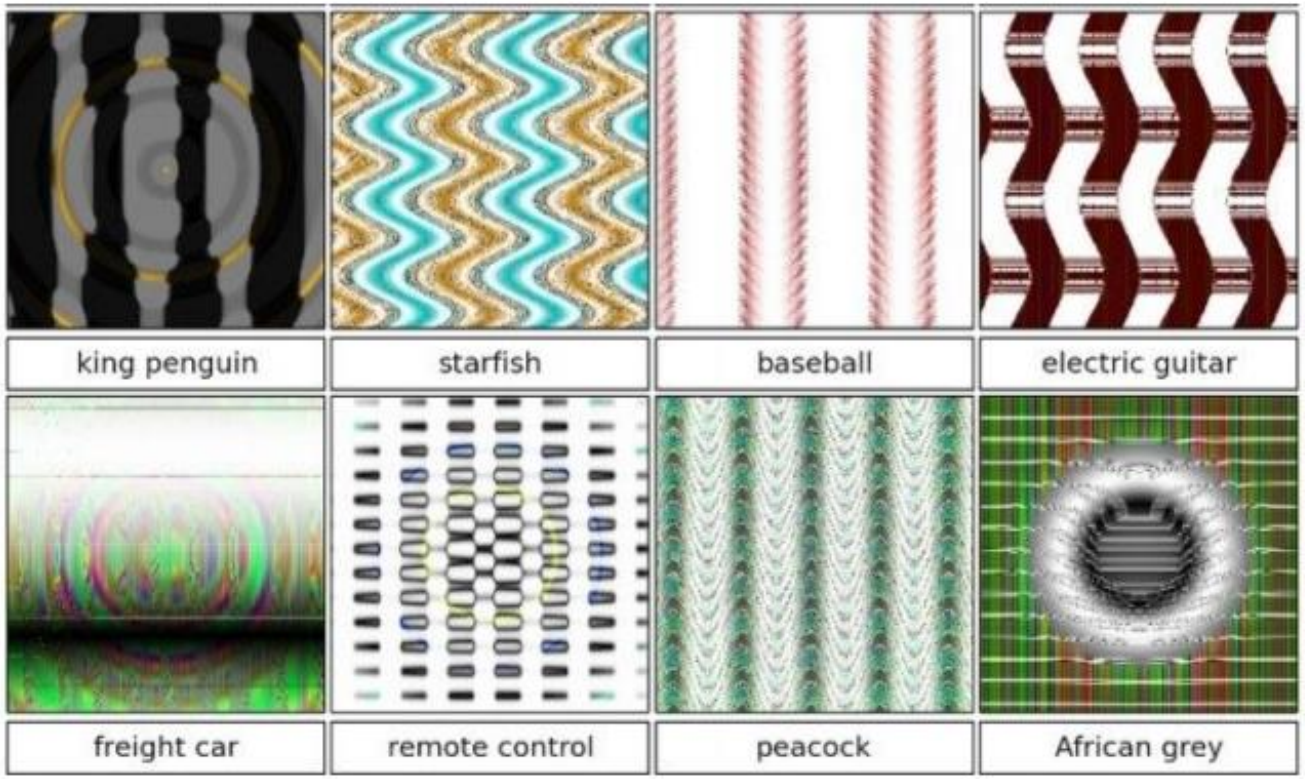


(b) After applying dropout.

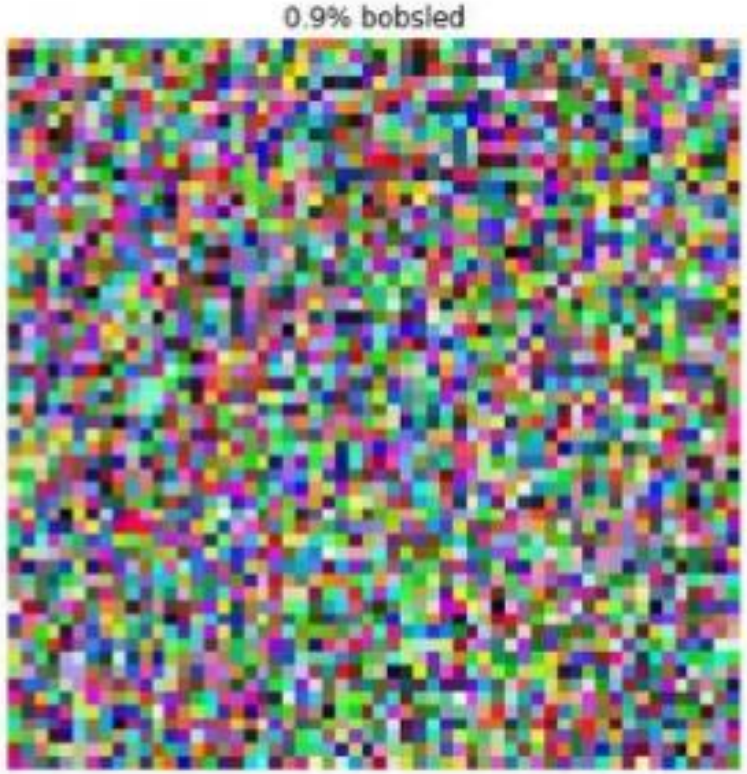
Fool your Conv-net

[Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images Nguyen, Yosinski, Clune, 2014]

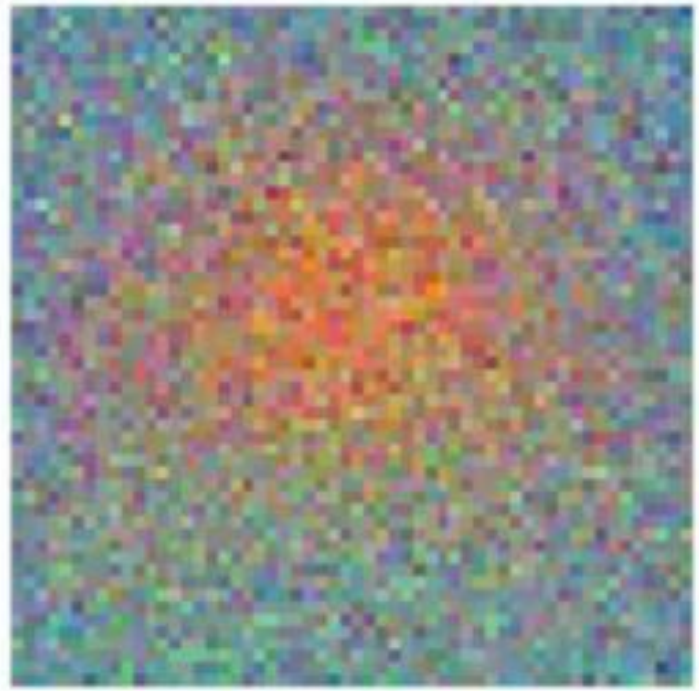
>99.6% confidences



mix in a tiny bit of
Goldfish classifier weights



+



=

100% Goldfish

Lets fool a binary linear classifier:

x	2	-1	3	-2	2	2	1	-4	5	1
w	-1	-1	1	-1	1	-1	1	1	-1	1
adversarial x	1.5	-1.5	3.5	-2.5	2.5	1.5	1.5	-3.5	4.5	1.5

← input example

← weights

class 1 score before:

-2 + 1 + 3 + 2 + 2 - 2 + 1 - 4 - 5 + 1 = -3

=> probability of class 1 is 1/(1+e^(-(-3))) = 0.0474

-1.5+1.5+3.5+2.5+2.5-1.5+1.5-3.5-4.5+1.5 = 2

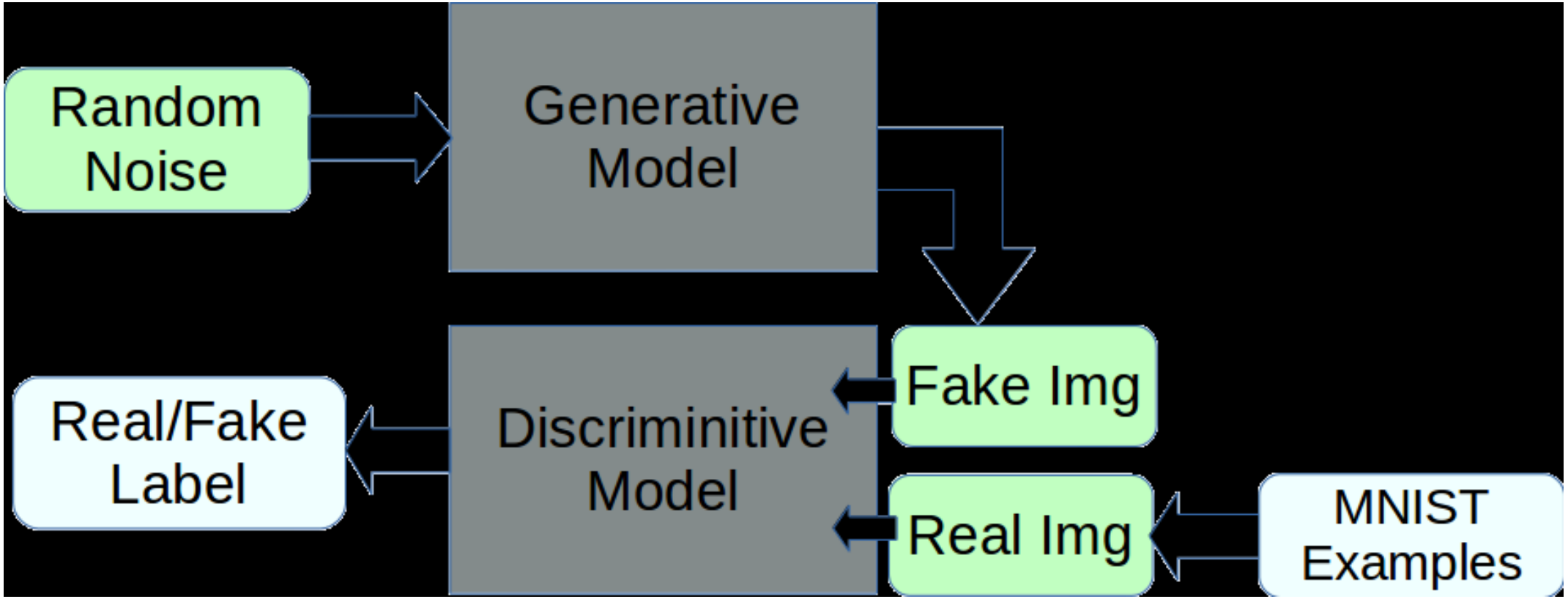
=> probability of class 1 is now 1/(1+e^(-(2))) = 0.88

i.e. we improved the class 1 probability from 5% to 88%

This was only with 10 input dimensions. A 224x224 input image has 150,528.

(It's significantly easier with more numbers, need smaller nudge for each)

Solution： The GAN framework



提出问题比解决问题更重要

提出的问题：

- Image caption
- Attention
- Generating images

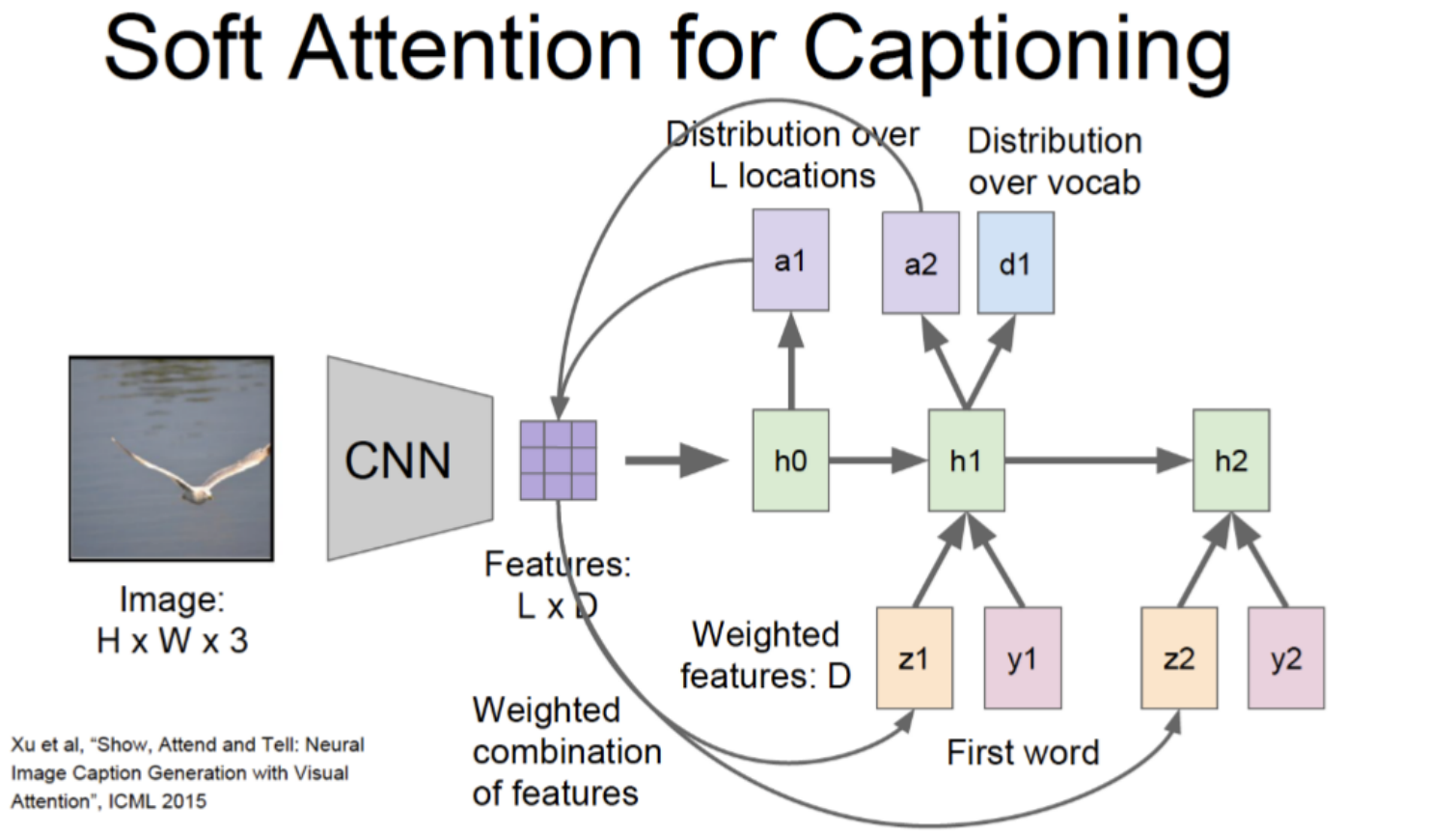
- 解决方案：
- 。。。总会有的

我们有什么 vs 我们想做什么

- Instance representation
- Temporal relationship

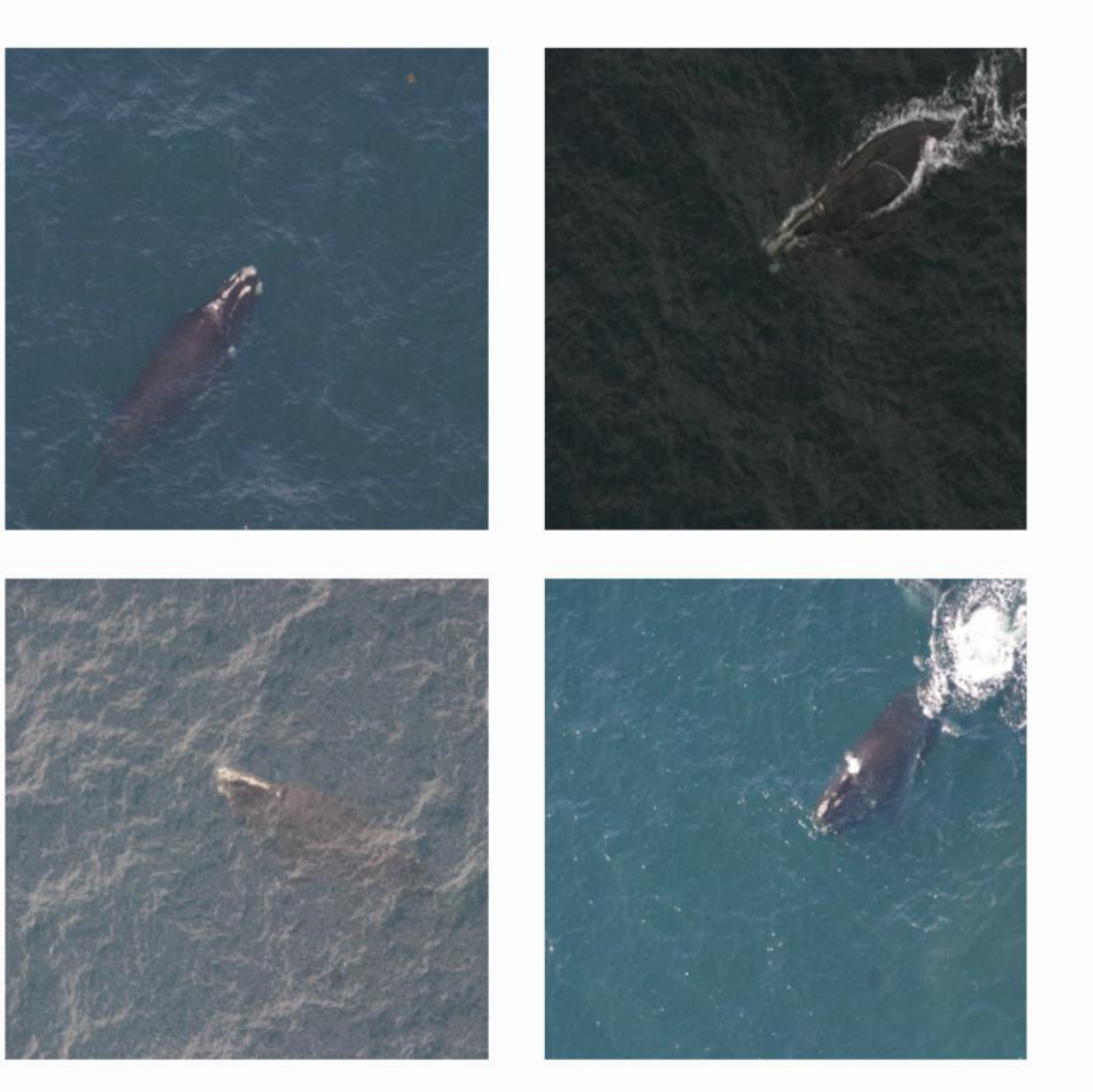
- Regression
- Classification

一个例子

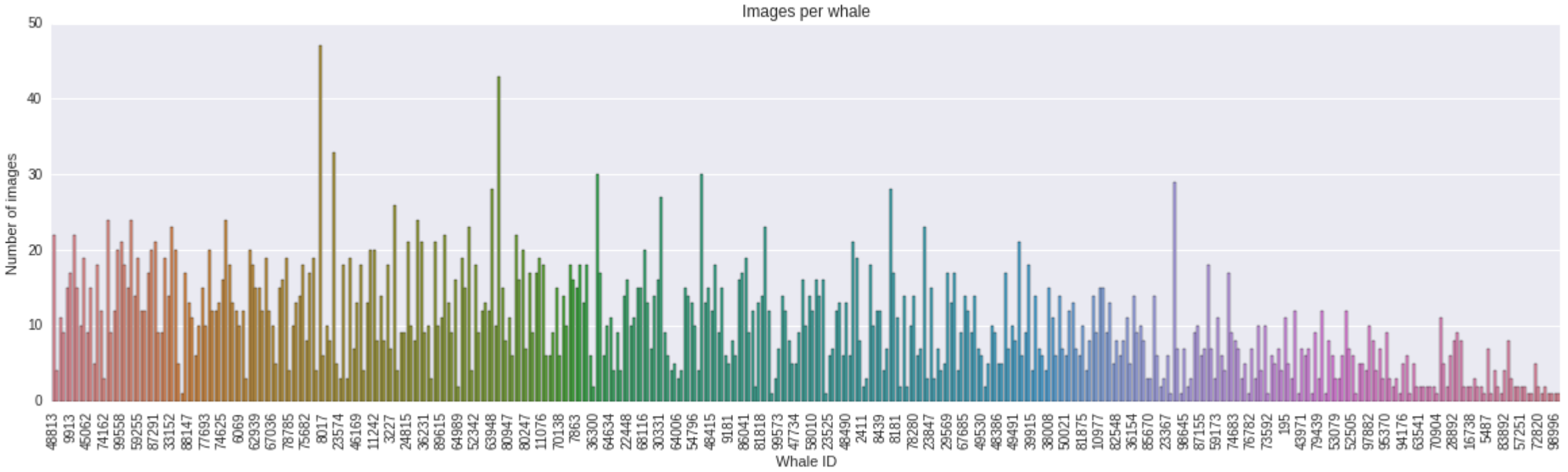

Soft Attention for Captioning
Distribution over
L locations
Distribution
over vocab
a1
a2
d1
CNN
h0
h1
h2
Image:
H x W x 3
Features:
L x D
Weighted
features: D
z1
y1
z2
y2
Weighted
combination
of features
First word
Xu et al, "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention", ICML 2015

Now， Kaggle Whale Challenge

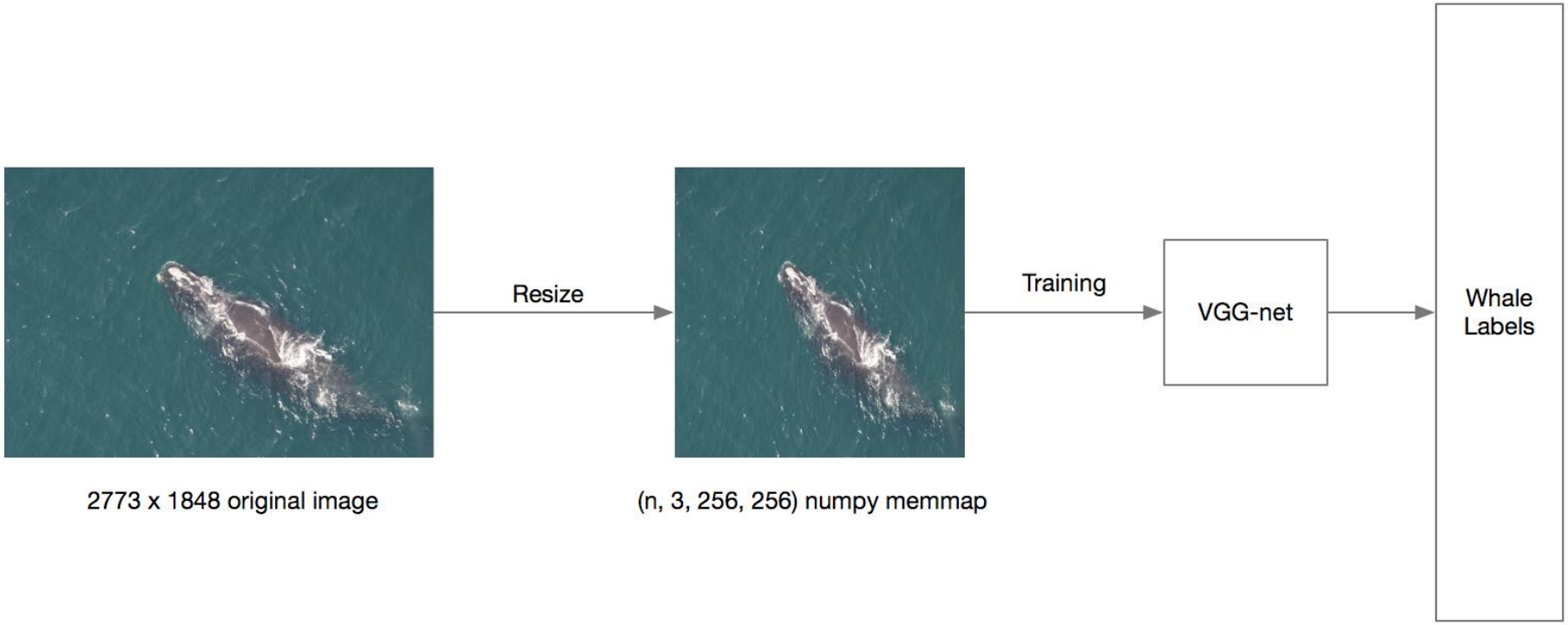
- Task

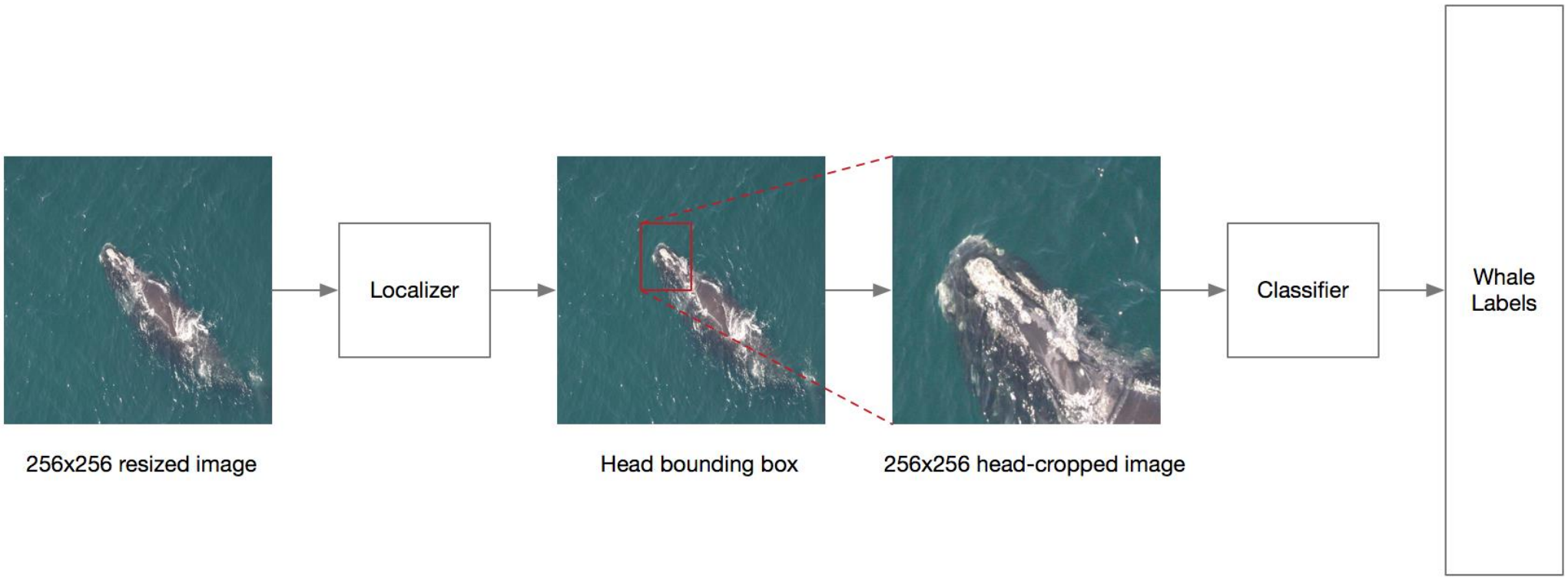


- 4237 images for 427 right whales



Baseline

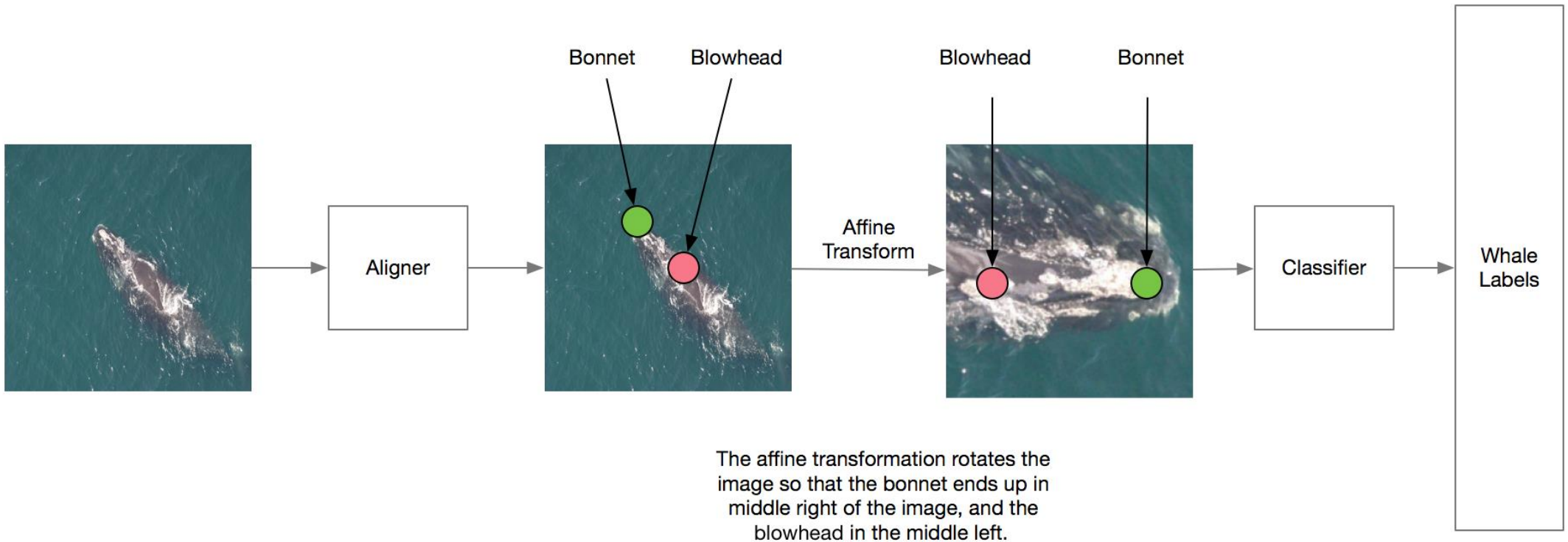

Resize
Training
VGG-net
Whale Labels
2773 x 1848 original image
(n, 3, 256, 256) numpy memmap


Localizer
Classifier
Whale Labels
256x256 resized image
Head bounding box
256x256 head-cropped image

How to get the bounding box

- 我们会什么？
- 分类 回归

- 我们想做什么？
- 获得bounding boxes



The affine transformation rotates the image so that the bonnet ends up in middle right of the image, and the blowhead in the middle left.

Second Example

车型识别

What makes AI great again

1. Compute (the obvious one: Moore's Law, GPUs, ASICs),
2. Data (in a nice form, not just out there somewhere on the internet - e.g. ImageNet),
3. Algorithms (research and ideas, e.g. backprop, CNN, LSTM), and
4. Infrastructure (software under you - Linux, TCP/IP, Git, TensorFlow,Theano, etc.).

Project codes

- Included in the attachment