Copyright Notice

deeplearning.ai

Error Analysis

Carrying out error analysis

Look at dev examples to evaluate ideas

90% accuracy

→ 10% error

Should you try to make your cat classifier do better on dogs?

Error analysis:

5-10 min

"ceiling"

- Get ~100 mislabeled dev set examples.
- Count up how many are dogs.

5% 10%

5/100 9.5%

50% 10%

50/100 5%

Andrew Ng

Evaluate multiple ideas in parallel

Ideas for cat detection:

- Fix pictures of dogs being recognized as cats
- Fix great cats (lions, panthers, etc..) being misrecognized
- Improve performance on blurry images

Image	Dog	Great Cats	Blurry	Instagram	Comments
1	✓			✓	Pitbull
2			✓	✓	
3		✓	✓		Rainy day at zoo
⋮	⋮	⋮	⋮		
% of total	8%	43%	61%	12%	

Andrew Ng

deeplearning.ai

Error Analysis

Cleaning up Incorrectly labeled data

Incorrectly labeled examples

x							
y	1	0	1	1	0	1	1

Training set.

DL algorithms are quite robust to random errors in the training set.

Systematic errors

Andrew Ng

Error analysis

Image	Dog	Great Cat	Blurry	Incorrectly labeled	Comments
...					
98				✓	Labeler missed cat in background
99		✓			
100				✓	Drawing of a cat; Not a real cat.
% of total	8%	43%	61%	6%	

Overall dev set error .... 10% | 2%

Errors due incorrect labels .... 0.6% | 0.6%

Errors due to other causes .... 9.4% | 1.4%

2.1% 1.9%

Goal of dev set is to help you select between two classifiers A & B.

Andrew Ng

Correcting incorrect dev/test set examples

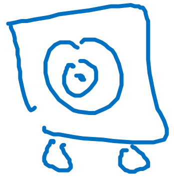
- Apply same process to your dev and test sets to make sure they continue to come from the same distribution
- Consider examining examples your algorithm got 98% right as well as ones it got wrong. 2%
- Train and dev/test data may now come from slightly different distributions.

deeplearning.ai

Error Analysis

Build your first system quickly, then iterate

Speech recognition example



- Noisy background
  - Café noise
  - Car noise
- Accent
- Far fro
- Young
- Stutter
- …

Guideline:

Build your first system quickly, then iterate

- Set up dev/test set and metric
- Build initial system quickly
- Use Bias/Variance analysis & Error analysis to prioritize next steps.

deeplearning.ai

Mismatched training and dev/test data

Training and testing on different distributions

Cat app example

care about this

Data from webpages

Data from mobile app

$\approx 200{,}000$ → 210,000 ← $\approx 10{,}000$

shuffle

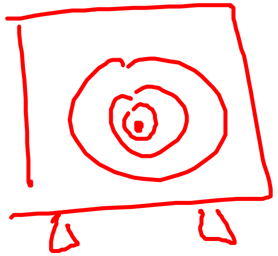
Option 1: | train | dev | test |

205,000

2,500

2,500

$\frac{200K}{210K}$



2381 - web

119 - mobile app

Option 2:

train

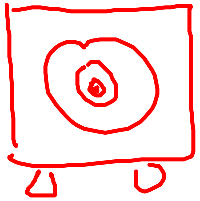
| web | dev | test |

train: 205,000

app

app 2500

app 2500



Andrew Ng

Speech recognition example

Speech activated rearview mirror

Training

Purchased data x, y

Smart speaker control

Voice keyboard

...

500,000 utterances

Dev/test

Speech activated rearview mirror

→ 20,000

10K 5K 5K

train D T

500K

D T

510K

10K mirror 5K 5K

It is often better to take out some data from the randomly shuffled dataset and to not train model on it. use rest of the data to train the model

Mismatched training and dev/test data

deeplearning.ai

Bias and Variance with mismatched data distributions

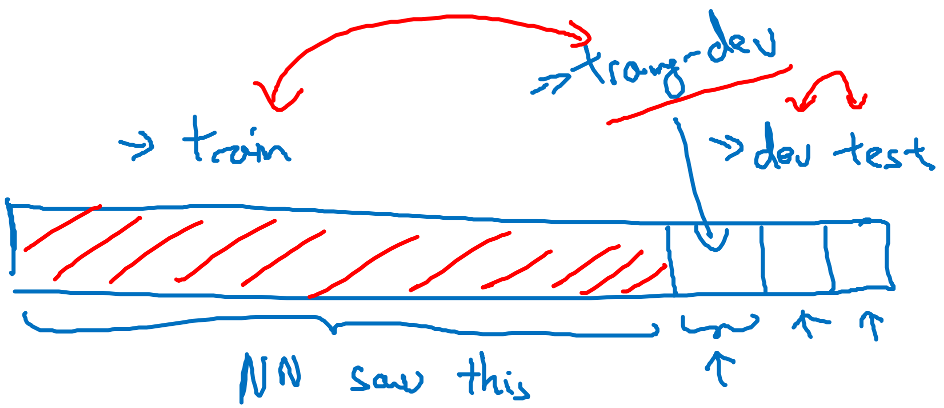
Cat classifier example

Assume humans get $\approx 0\%$ error.

Training error 1%

Dev error 10%

9%



Training error	1%	1%
Training-dev error	9%	1.5%
Dev error	10%	10%

variance

data mismatch

Variance

Human error	0%	
Training error	10%	10%
Training-dev error	11%	11%
Dev error	12%	20%

Avoidable bias

Avoidable bias

variance

Data mismatch

Bias

Bias + Data mismatch

Training-dev set: Same distribution as training set, but not used for training

Bias/variance on mismatched training and dev/test sets

Human level	4%		4%
Training set error	7%	↕ avoidable bias	7%
Training-dev set error	10%	↕ variance	10%
→ Dev error	12%	↕ data mismatch	6%
→ Test error	12%	↕ degree of overfitting to dev set.	6%

More general formulation

Rearview mirror

	General speech recognition	Rearview mirror speech data
Human level	"Human level" 4%	6%
Error on examples trained on	"Training error" 7%	6%
Error on example not trained on	"Training-dev error" 10%	"Dev/Test error" 6%

Avoidable bias

Variance

data mismatch

Mismatched training and dev/test data

deeplearning.ai

Addressing data mismatch

Addressing data mismatch

- Carry out manual error analysis to try to understand difference between training and dev/test sets

  E.g. noisy – car noise   street numbers

- Make training data more similar; or collect more data similar to dev/test sets

  E.g. Simulate noisy in-car data

Andrew Ng

Artificial data synthesis

+ =

"The quick brown fox jumps over the lazy dog."

10,000 hours

Car noise

1 hour of car noise

Overfit to 1 hour of car noise

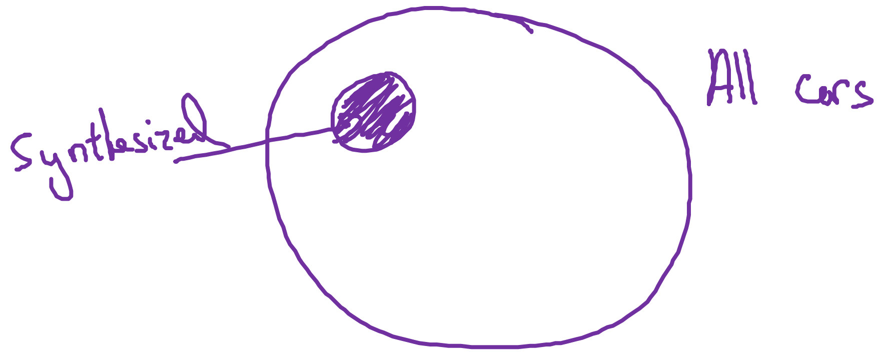
10,000 hours

Synthesized in-car audio

Artificial data synthesis

Car recognition:

≈ 20 cars



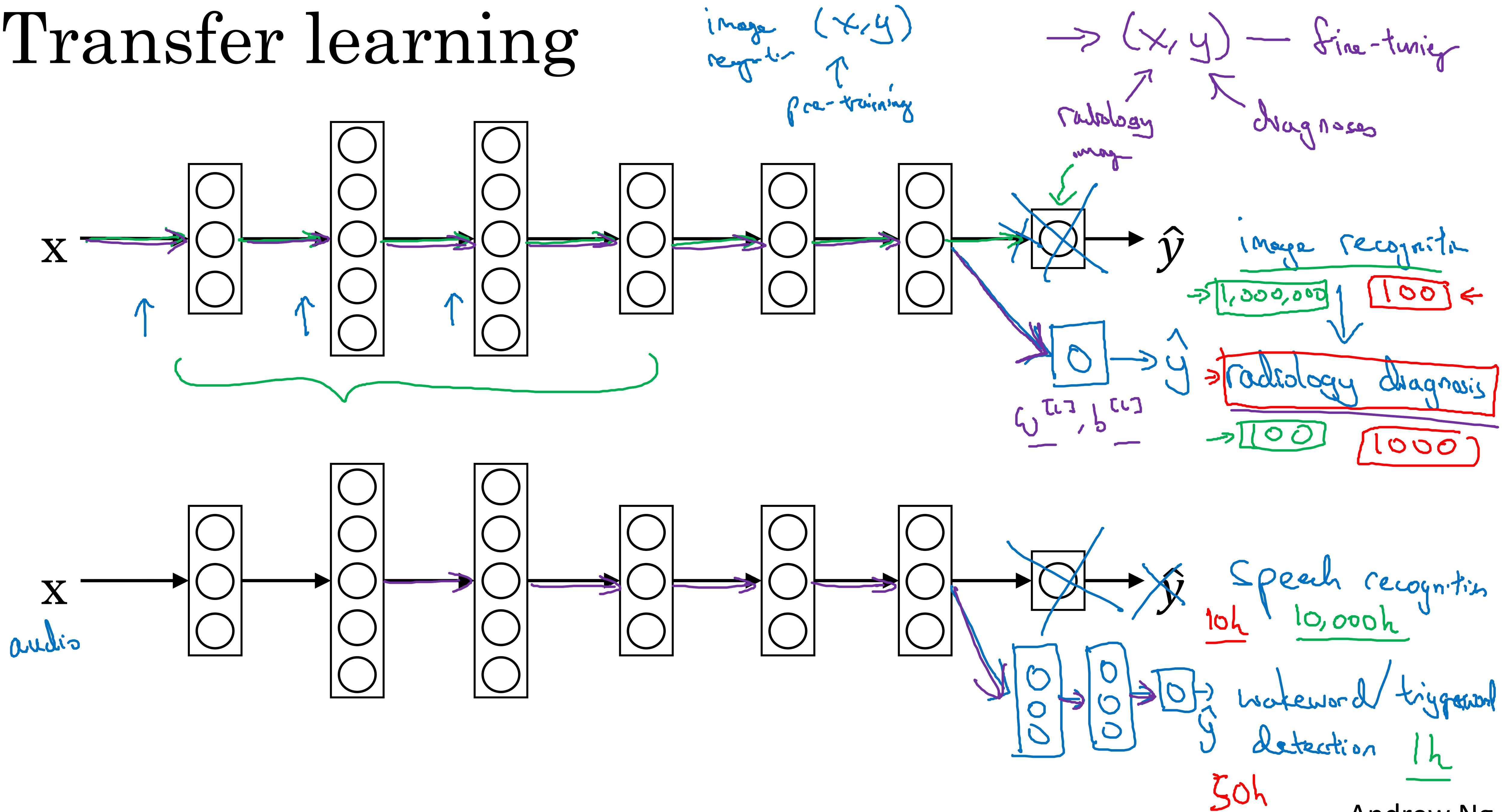
Andrew Ng

deeplearning.ai

Learning from multiple tasks

Transfer learning

Transfer learning



When transfer learning makes sense

Transfer from A → B

- Task A and B have the same input x.
- You have a lot more data for Task A than Task B.
- Low level features from A could be helpful for learning B.

deeplearning.ai

Learning from multiple tasks

Multi-task learning

Simplified autonomous driving example

$x^{(i)}$

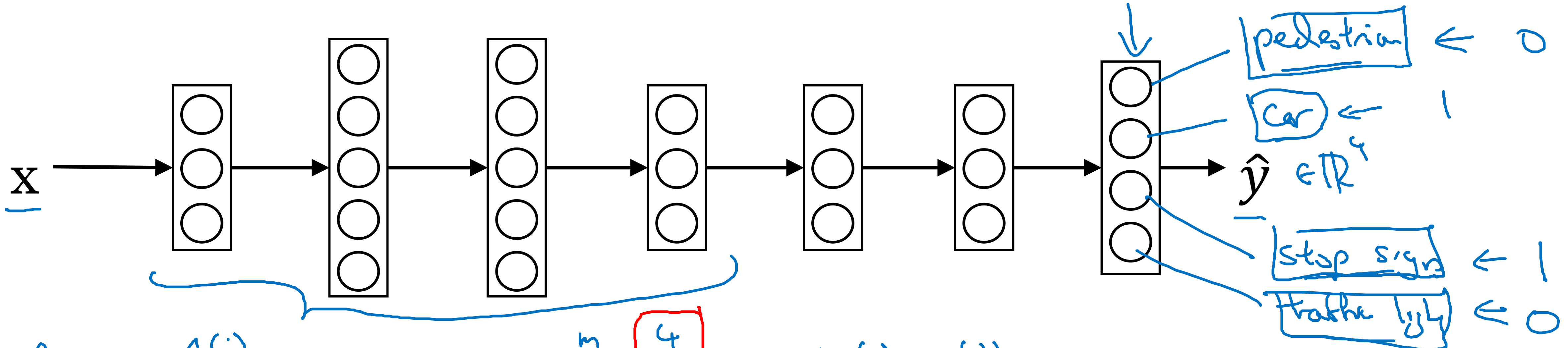
$y^{(i)}$ $(4,1)$

	$y^{(i)}$
Pedestrians	0
Cars	1
Stop signs	1
Traffic lights	0
⋮	⋮

$$Y = \begin{bmatrix} | & | & | & & | \\ y^{(1)} & y^{(2)} & y^{(3)}, & \cdots, & y^{(m)} \\ | & | & | & & | \end{bmatrix}$$

$(4, m)$

Neural network architecture



Loss: $\underset{(4,1)}{\hat{y}^{(i)}}$ $\rightarrow \frac{1}{m}\sum_{i=1}^{m}\boxed{\sum_{j=1}^{4}} \mathcal{L}(\hat{y}_j^{(i)}, y_j^{(i)})$

Sum only over value of j with 0/1 label.

→ Usual logistic loss

$-y_j^{(i)}\log\hat{y}_j^{(i)} - (1-y_j^{(i)})\log(1-\hat{y}_j^{(i)})$

Unlike softmax regression:

One image can have multiple labels

Multi-task learning ←

$$Y = \begin{bmatrix} 1 & \boxed{1} & & \boxed{0} & & ? & \\ 0 & \boxed{1} & \cdots & \boxed{1} & \cdots & 1 & \cdots \\ \boxed{?} & ? & & \boxed{1} & & ? & \\ ? & ? & & \boxed{0} & & ? & \end{bmatrix} \leftarrow$$

Andrew Ng

When multi-task learning makes sense

- Training on a set of tasks that could benefit from having shared lower-level features.
- Usually: Amount of data you have for each task is quite similar.

A 1,000,000

B 1,000

A_1 1,000

A_2 1,000

$\vdots$

A_{100} 1,000

99,000

- Can train a big enough neural network to do well on all the tasks.

deeplearning.ai

End-to-end deep learning

What is end-to-end deep learning

What is end-to-end learning?

Speech recognition example

"cat"

x audio $\xrightarrow{\text{MFCC}}$ features $\xrightarrow{\text{ML}}$ Phonemes $\rightarrow$ words $\rightarrow$ transcript y

audio $\longrightarrow$ phonemes $\longrightarrow$ ... $\longrightarrow$

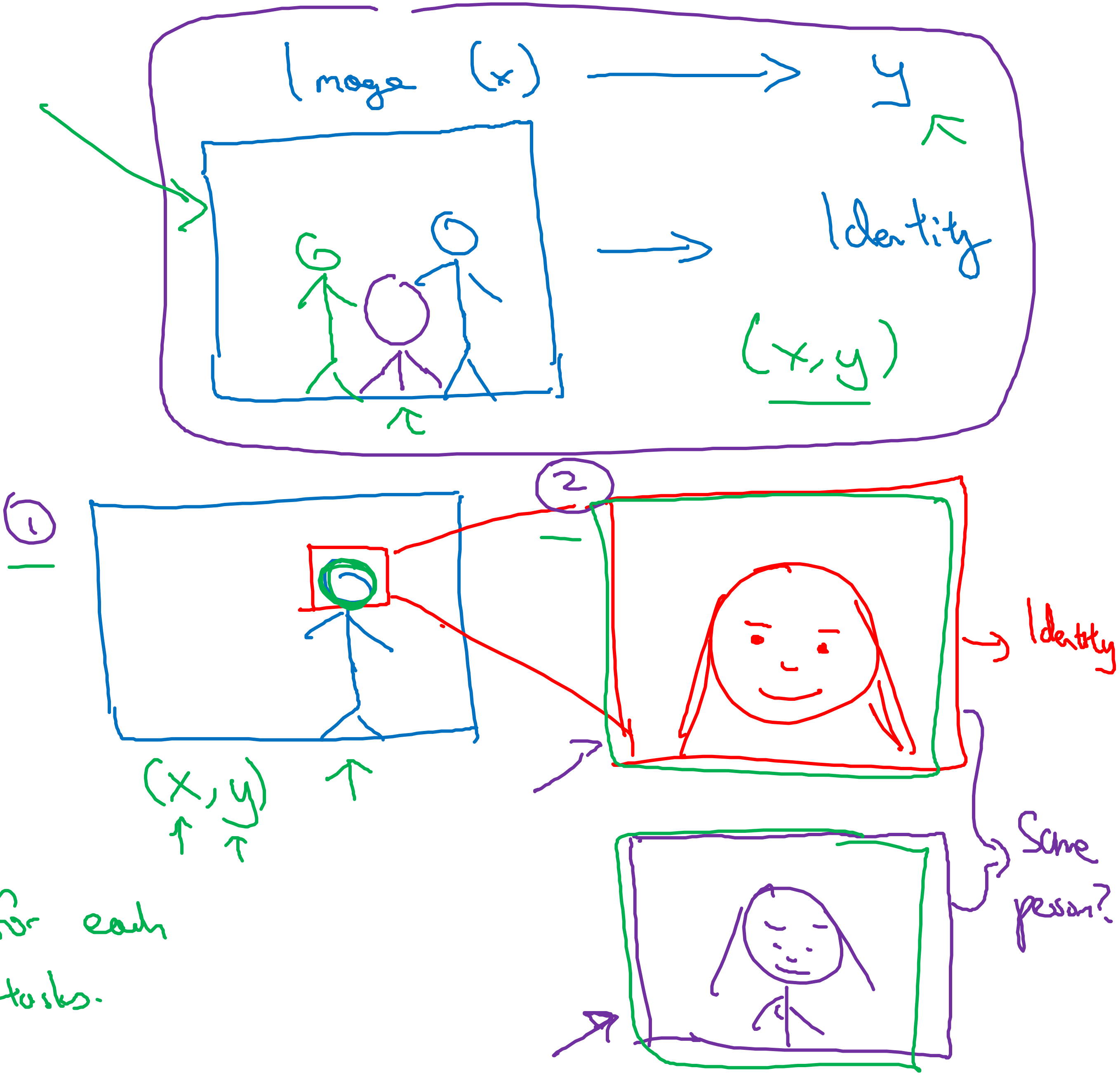
$\rightarrow$ audio $\longrightarrow$ transcript

3,000h

10,000h
⋮
100,000h

Face recognition

[Image courtesy of Baidu]



Have data for each of 2 subtasks.

More examples

Machine translation

(x, y)

English   French

English $\rightarrow$ text analysis $\rightarrow$ ... $\rightarrow$ French

English $\longrightarrow$ French

Estimating child's age:

Image $\xrightarrow{1}$ bones $\xrightarrow{2}$ age

Image $\longrightarrow$ age

Andrew Ng

End-to-end deep learning

deeplearning.ai

Whether to use end-to-end learning

Pros and cons of end-to-end deep learning

Pros:

- Let the data speak

$x \longrightarrow y$

"phonemes"

c a t

- Less hand-designing of components needed

$x - - - - - \rightarrow y$

input end

output end

Cons:

- May need large amount of data

$x \longrightarrow y$ (x, y)

- Excludes potentially useful hand-designed components

Data.

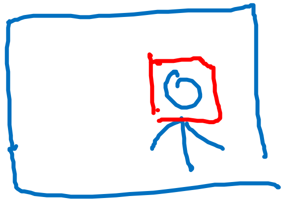
Hand-design.

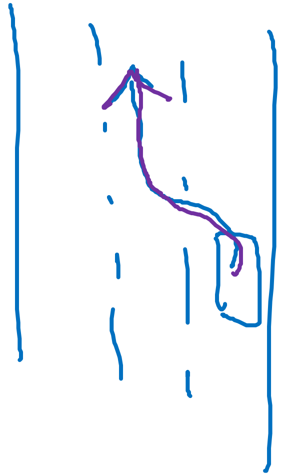
Applying end-to-end deep learning

Key question: Do you have sufficient data to learn a function of the complexity needed to map x to y?

X → y

→ age





image, radar, lidar $\xrightarrow{DL}$ cars, $\xrightarrow{DL}$ pedestrians → motion planning → route $\xrightarrow{\text{control}}$ steering

- Use DL to learn individual components
- Carefully choose X → Y depending on what tasks you can get data for.

→ image ⟶ steering

Andrew Ng