

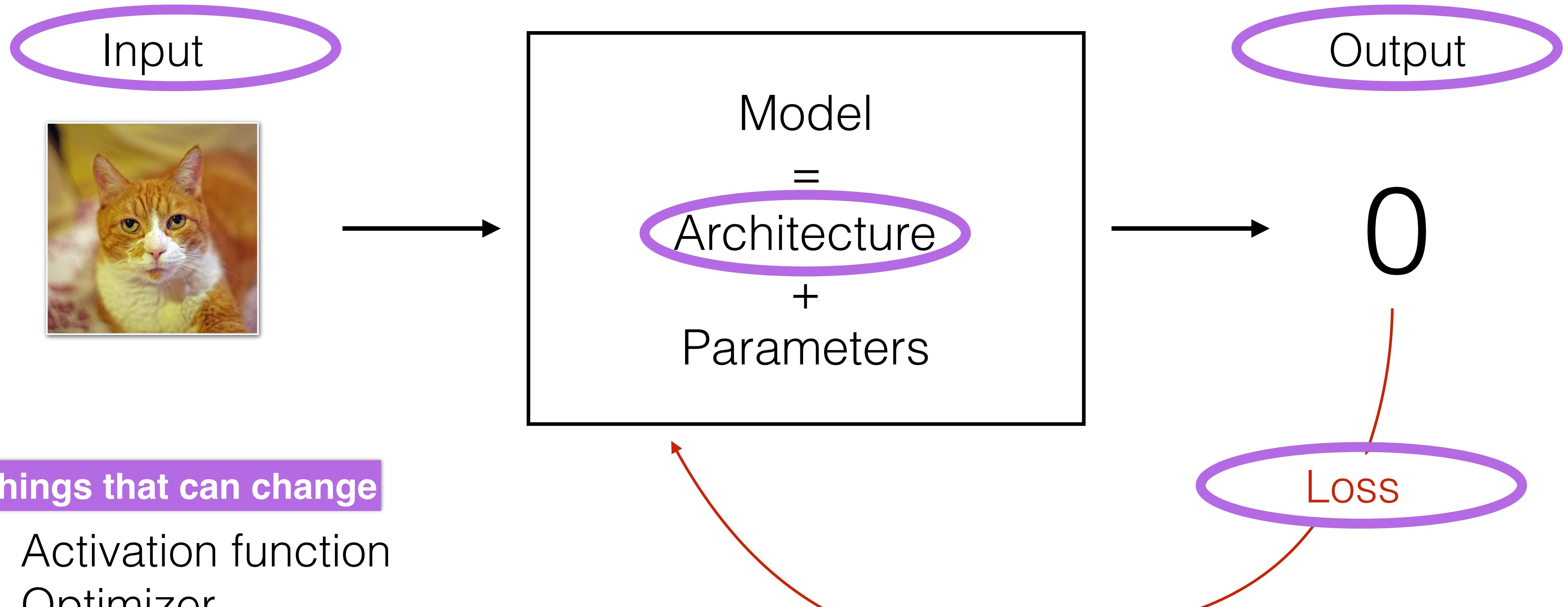
CS230: Lecture 3

Deep Learning Intuition

Kian Katanforoosh

Recap

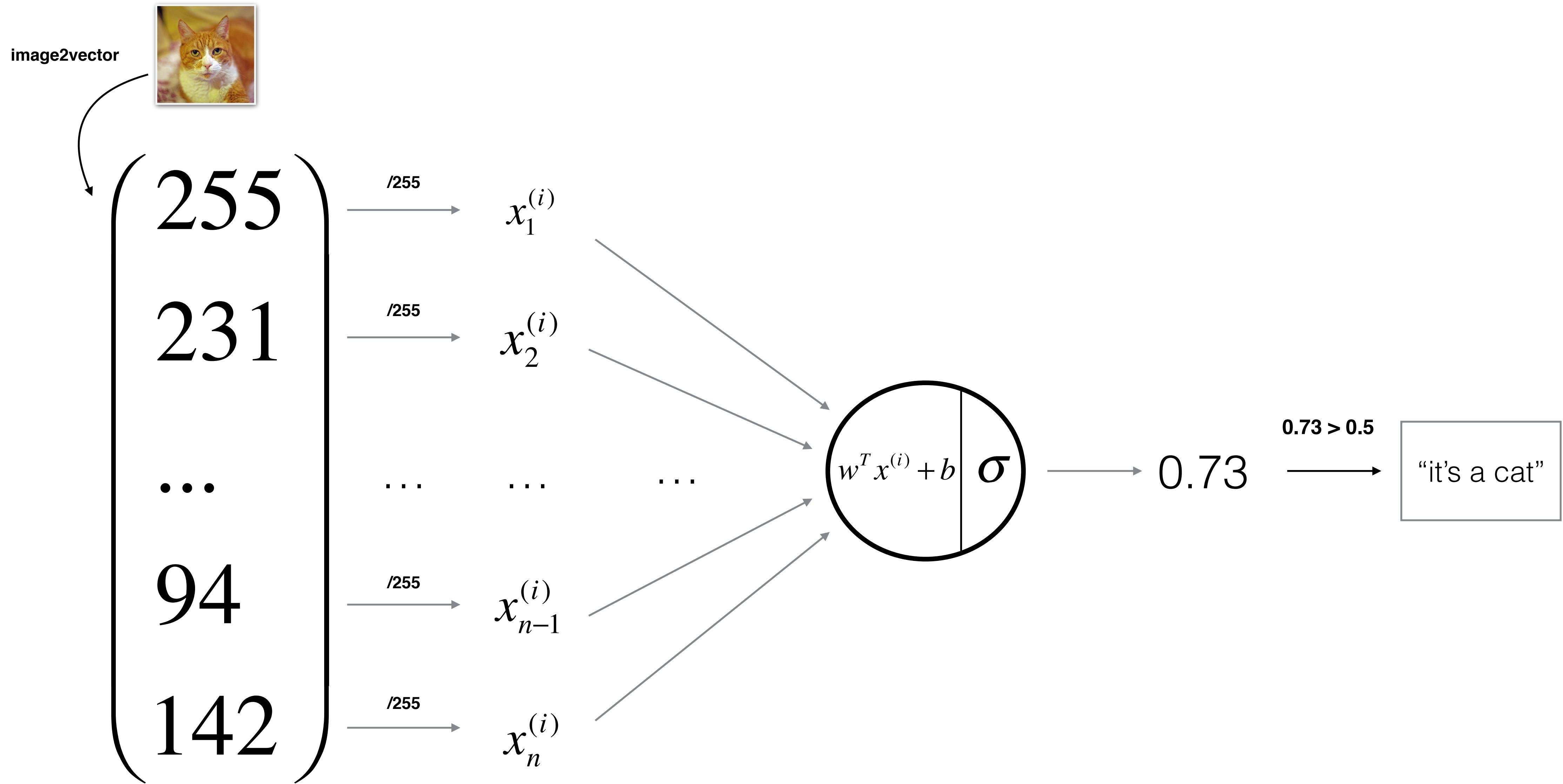
Learning Process



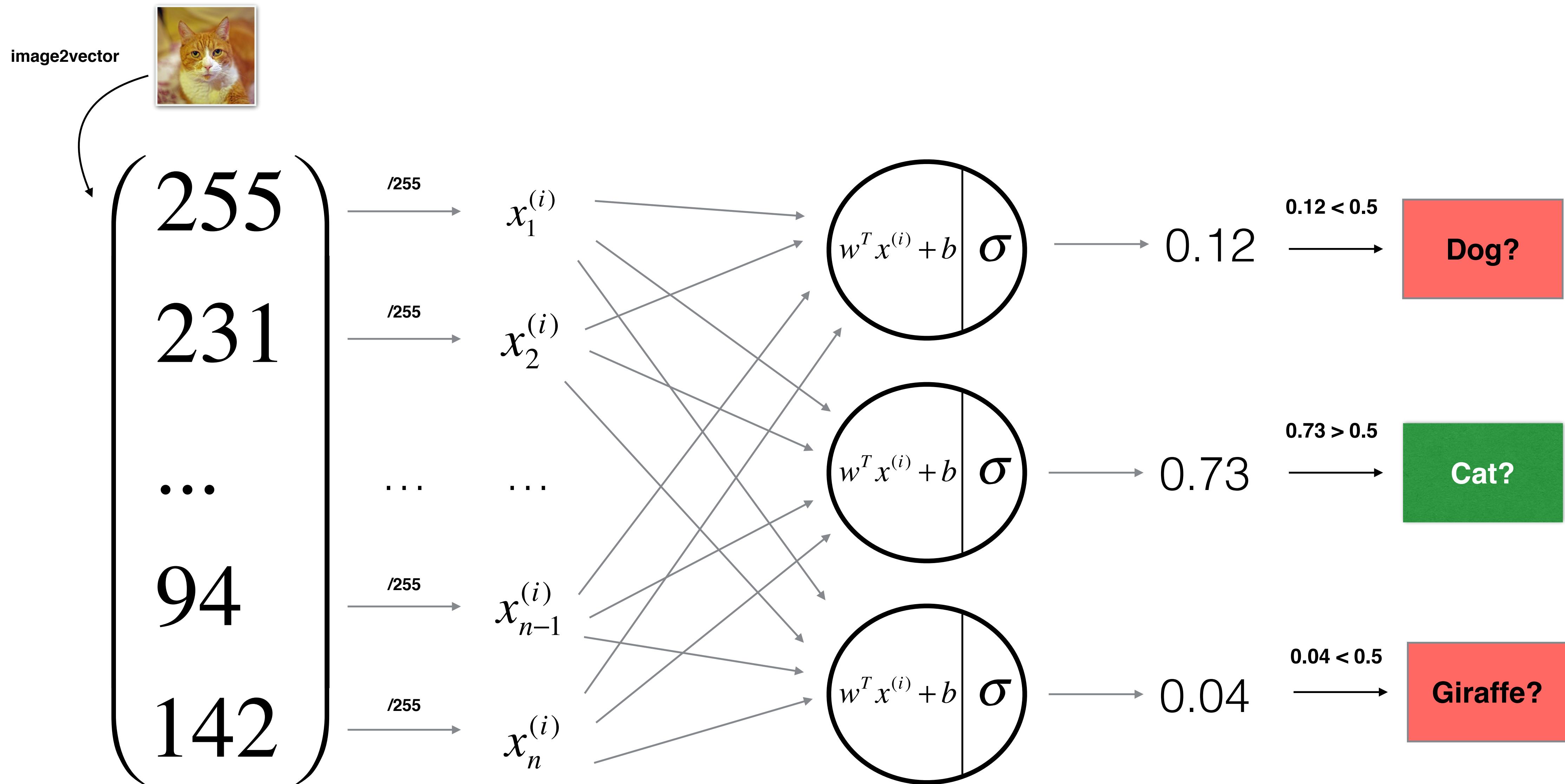
Things that can change

- Activation function
- Optimizer
- Hyperparameters
- ...

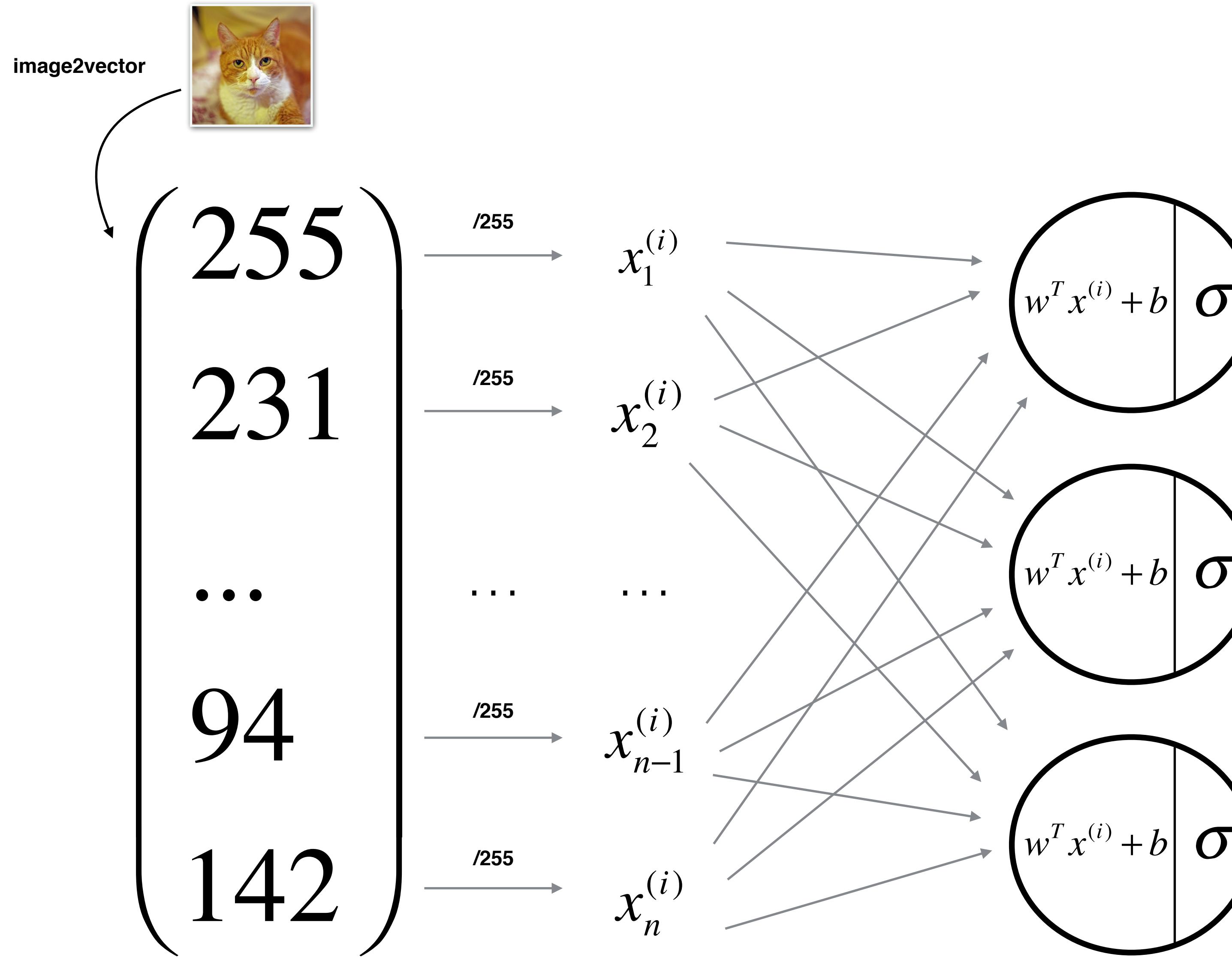
Logistic Regression as a Neural Network



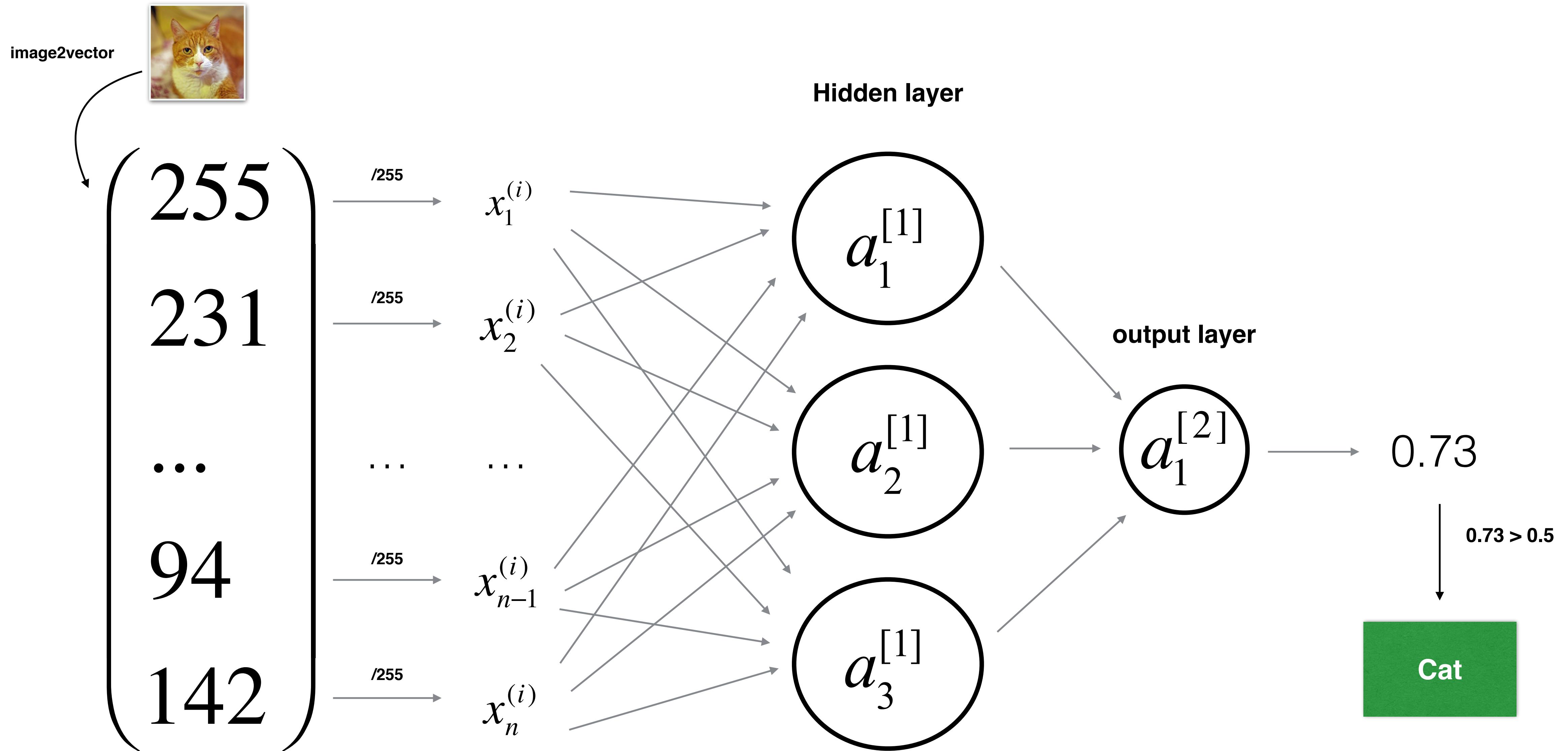
Multi-class



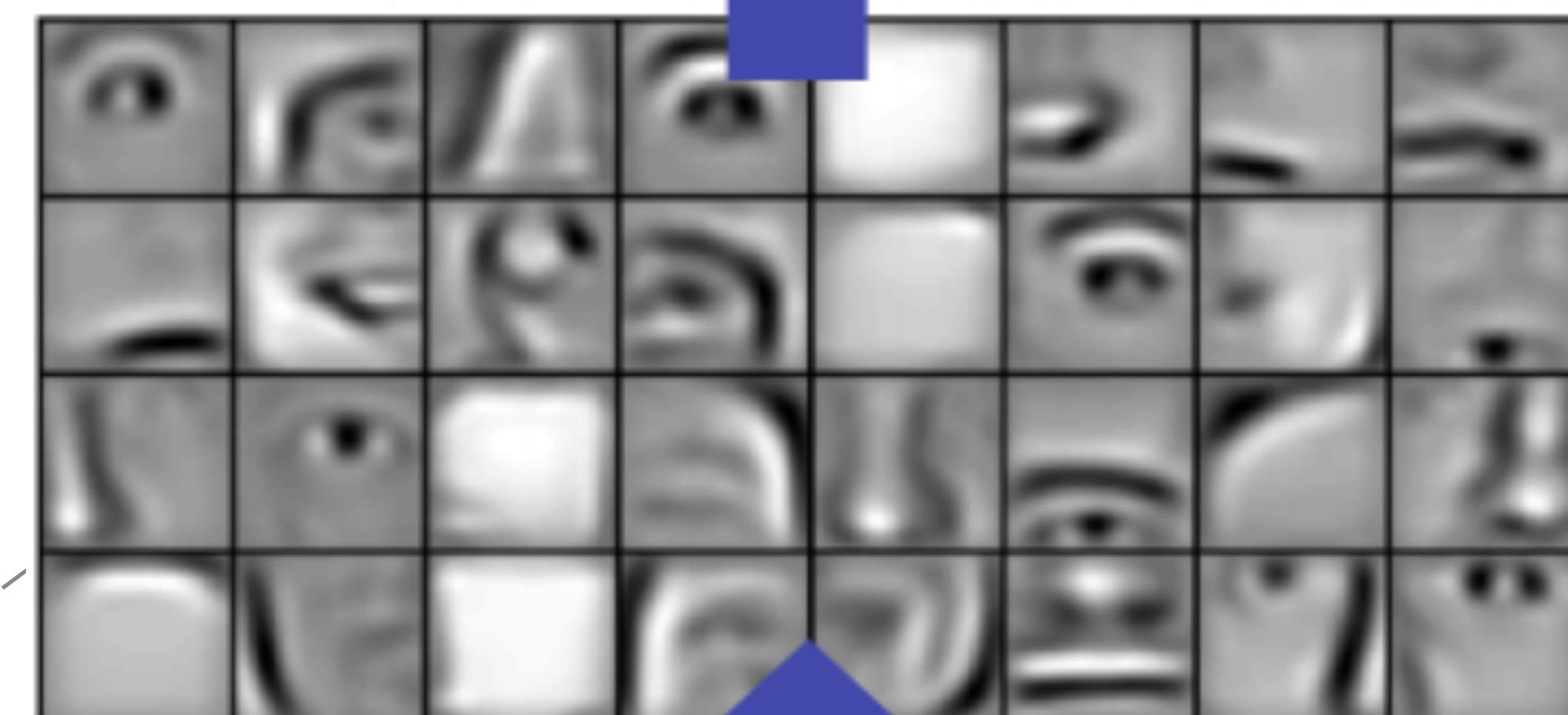
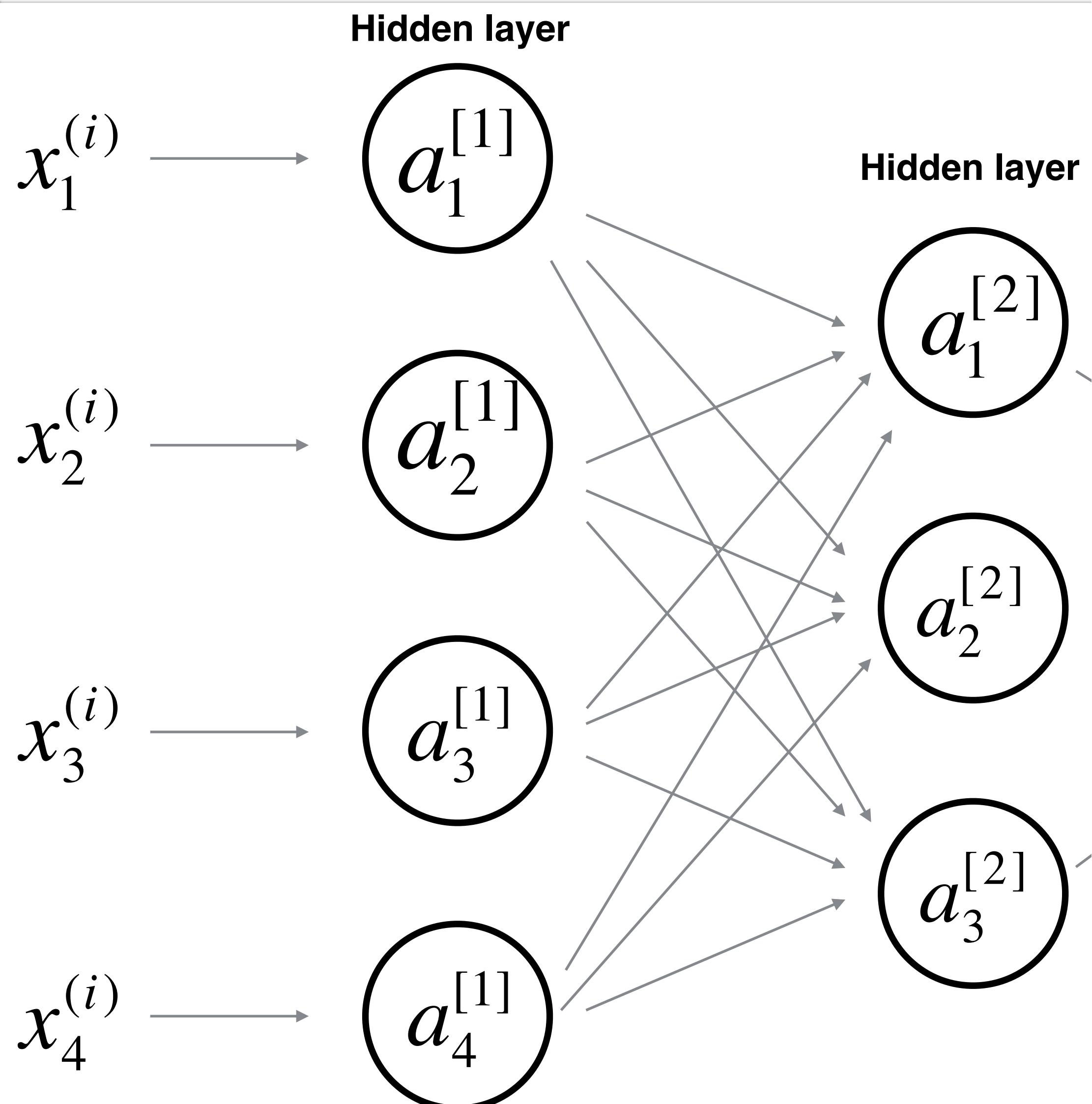
Neural Network (Multi-class)



Neural Network (1 hidden layer)



Deeper net



Technique called “encoding”

Let's build intuition on concrete applications

Today's outline

We will learn tips and tricks to:

- Analyze a problem from a deep learning approach
 - Choose an **architecture**
 - Choose a **loss** and a **training strategy**
- I. Day'n'Night classification
 - II. Face verification and recognition
 - III. Neural style transfer (Art generation)
 - IV. Trigger-word detection
 - V. Shipping model

Day'n'Night classification

Goal: Given an image, classify as taken “during the day” (0) or “during the night” (1)

1. Data?

10,000 images

Split? Bias?

2. Input?



Resolution?

(64, 64, 3)

3. Output?

y = 0 or y = 1

Last Activation?

sigmoid

4. Architecture ?

Shallow network should do the job pretty well

5. Loss?

$$L = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

Easy warm up

Face Verification

Goal: A school wants to use Face Verification for validating student IDs in facilities (dinning halls, gym, pool ...)

1. Data?

Picture of every student labelled with their name



Bertrand

2. Input?



Resolution?
(412, 412, 3)

3. Output?

$y = 1$ (it's you)
or
 $y = 0$ (it's not you)

Face Verification

Goal: A school wants to use Face Verification for validating student IDs in facilities (dinning halls, gym, pool ...)

4. What architecture?

Simple solution:



compute distance
pixel per pixel
if less than threshold
then $y=1$



database image

input image

Issues:

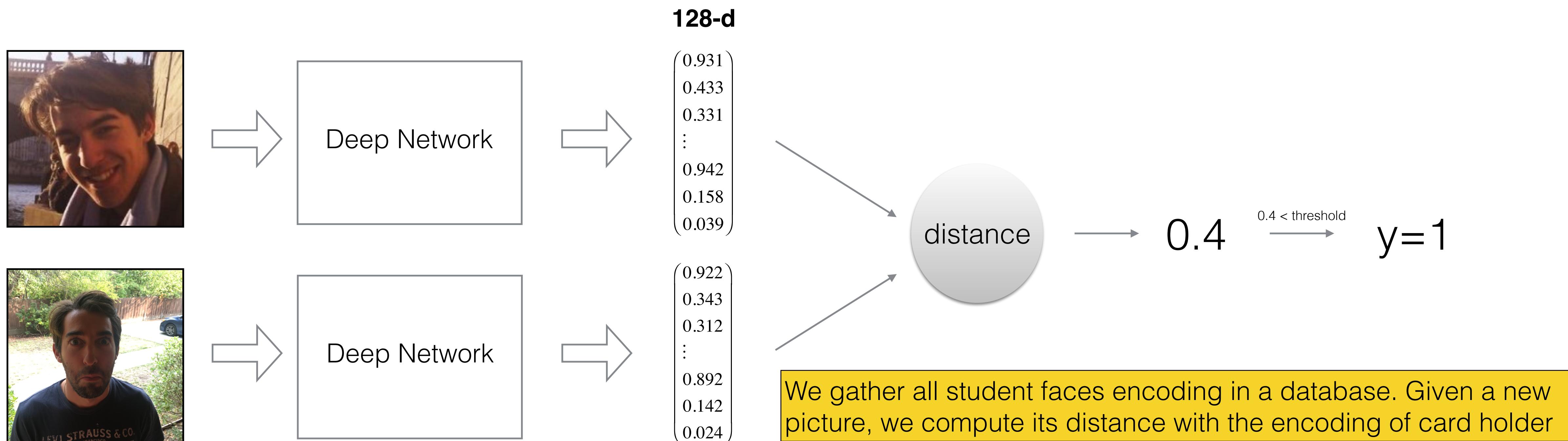
- Background lighting differences
- A person can wear make-up, grow a beard...
- ID photo can be outdated

Face Verification

Goal: A school wants to use Face Verification for validating student IDs in facilities (dinning halls, gym, pool ...)

4. What architecture?

Our solution: encode information about a picture in a vector



Face Recognition

Goal: A school wants to use Face Verification for validating student IDs in facilities (dinning hall, gym, pool ...)

4. Loss? Training?

We need more data so that our model understands how to encode:
Use public face datasets

What we really want:



similar encoding



different encoding

So let's generate triplets:



anchor



positive



negative

minimize encoding distance

maximize encoding distance

Face Recognition

What we really want:



similar encoding



different encoding

So let's generate triplets:



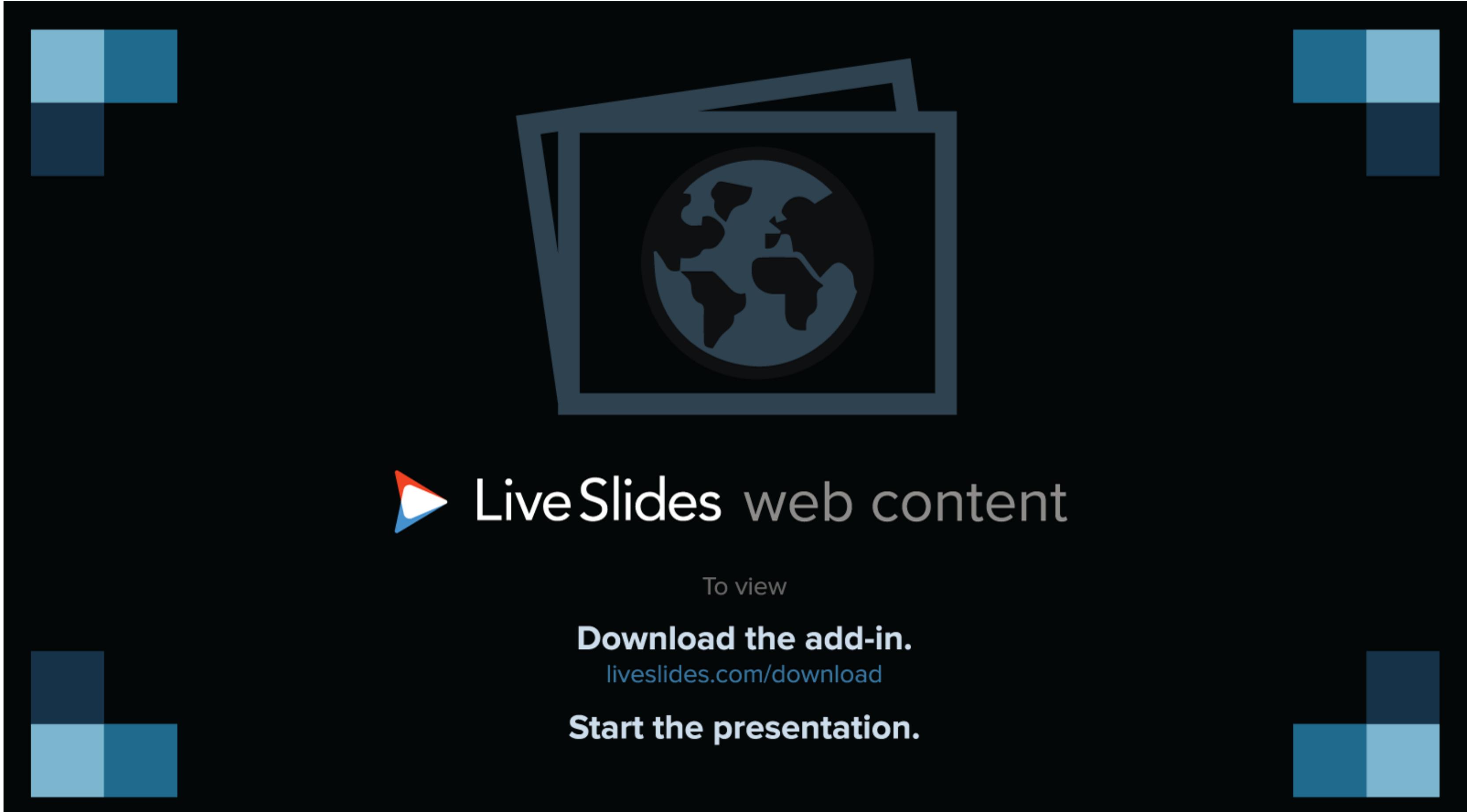
anchor

positive

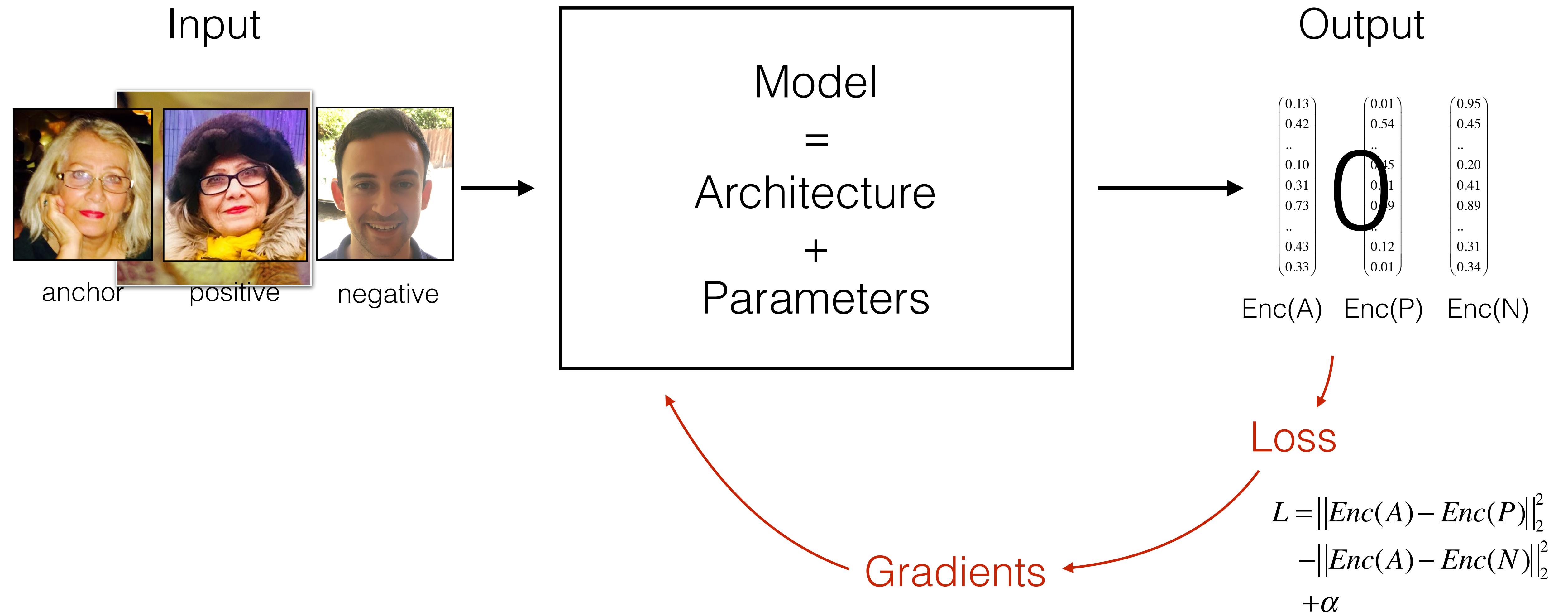
negative

minimize encoding distance

maximize encoding distance



Recap: Learning Process



Face Recognition

Goal: A school wants to use Face Identification for recognize students in facilities (dinning hall, gym, pool ...)

K-Nearest Neighbors

Goal: You want to use Face Clustering to group pictures of the same people on your smartphone

K-Means Algorithm

Maybe we need to detect the faces first?

Art generation (Neural Style Transfer)

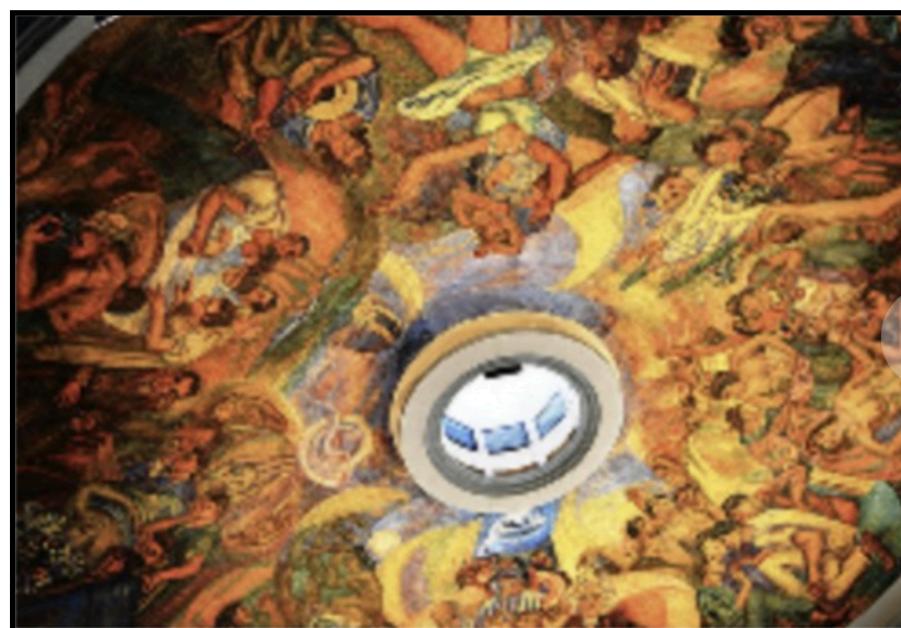
Goal: Given a picture, make it look beautiful

1. Data?

Let's say we have
any data

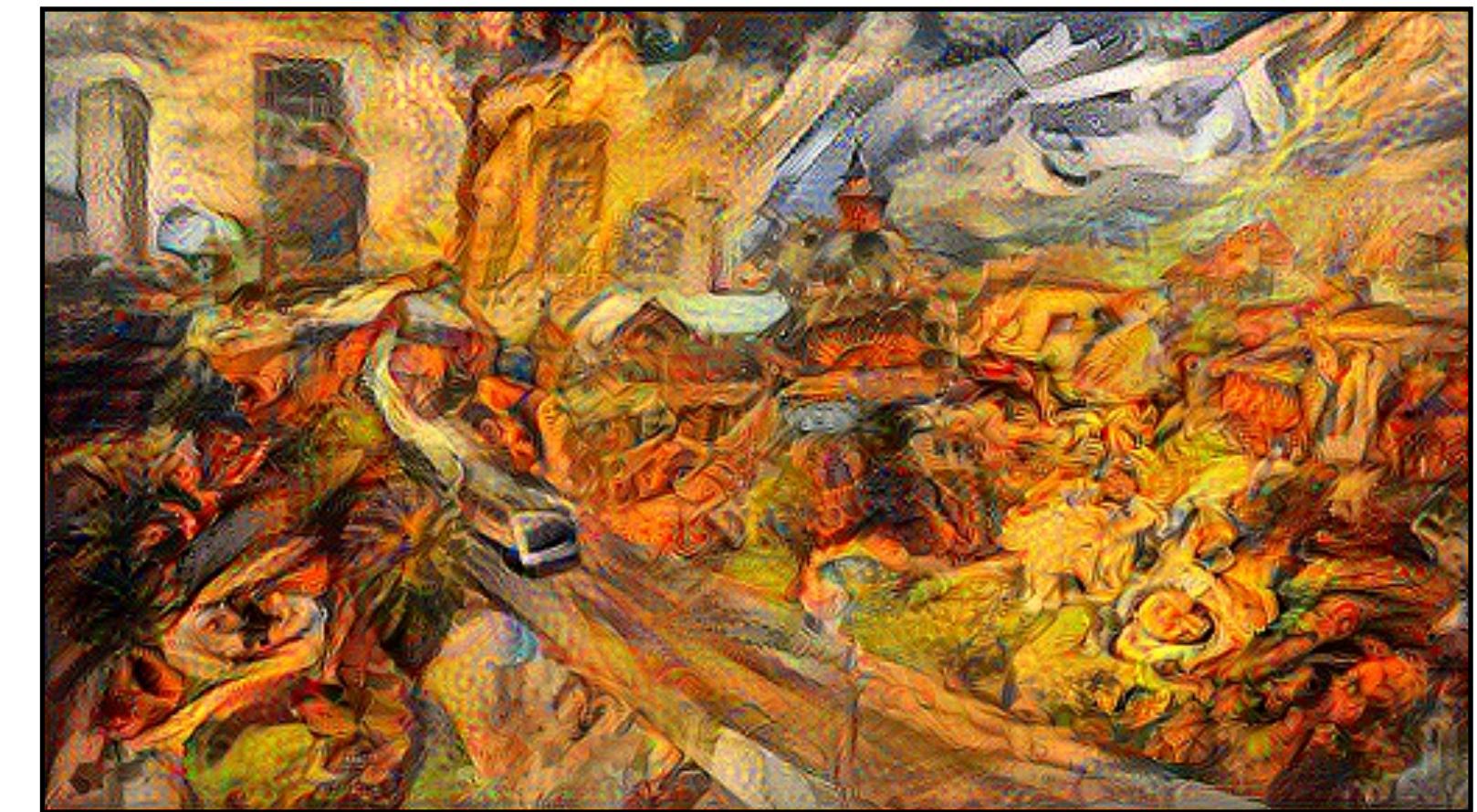


2. Input?



content
image

3. Output?



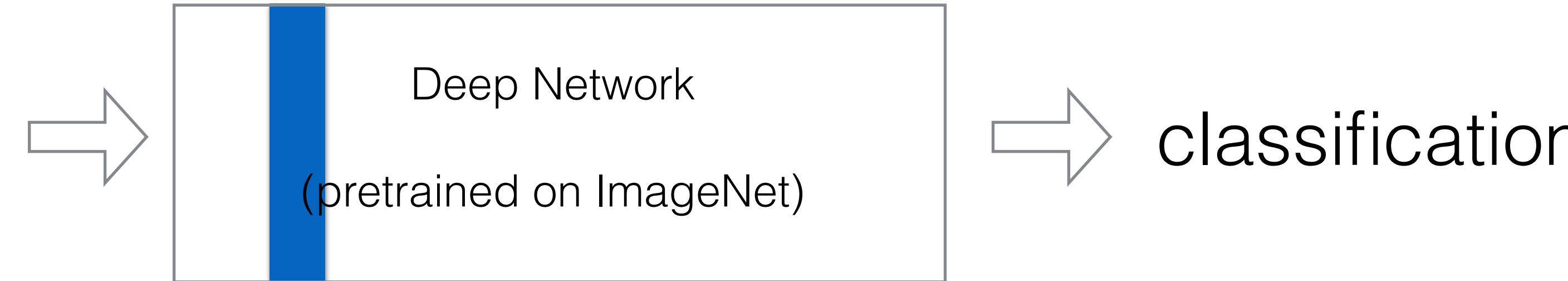
style
image

generated
image

Art generation (Neural Style Transfer)

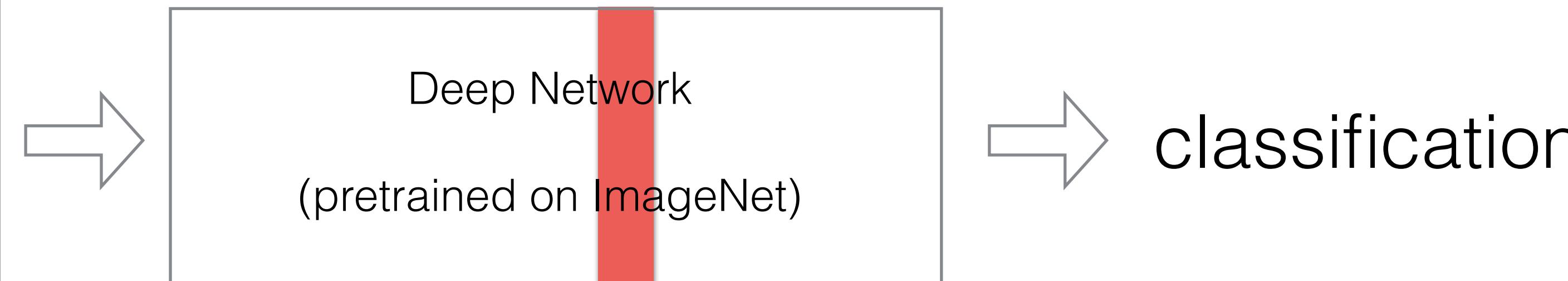
4. Architecture?

We use a **pre-trained model** because it **extracts important information** from images.



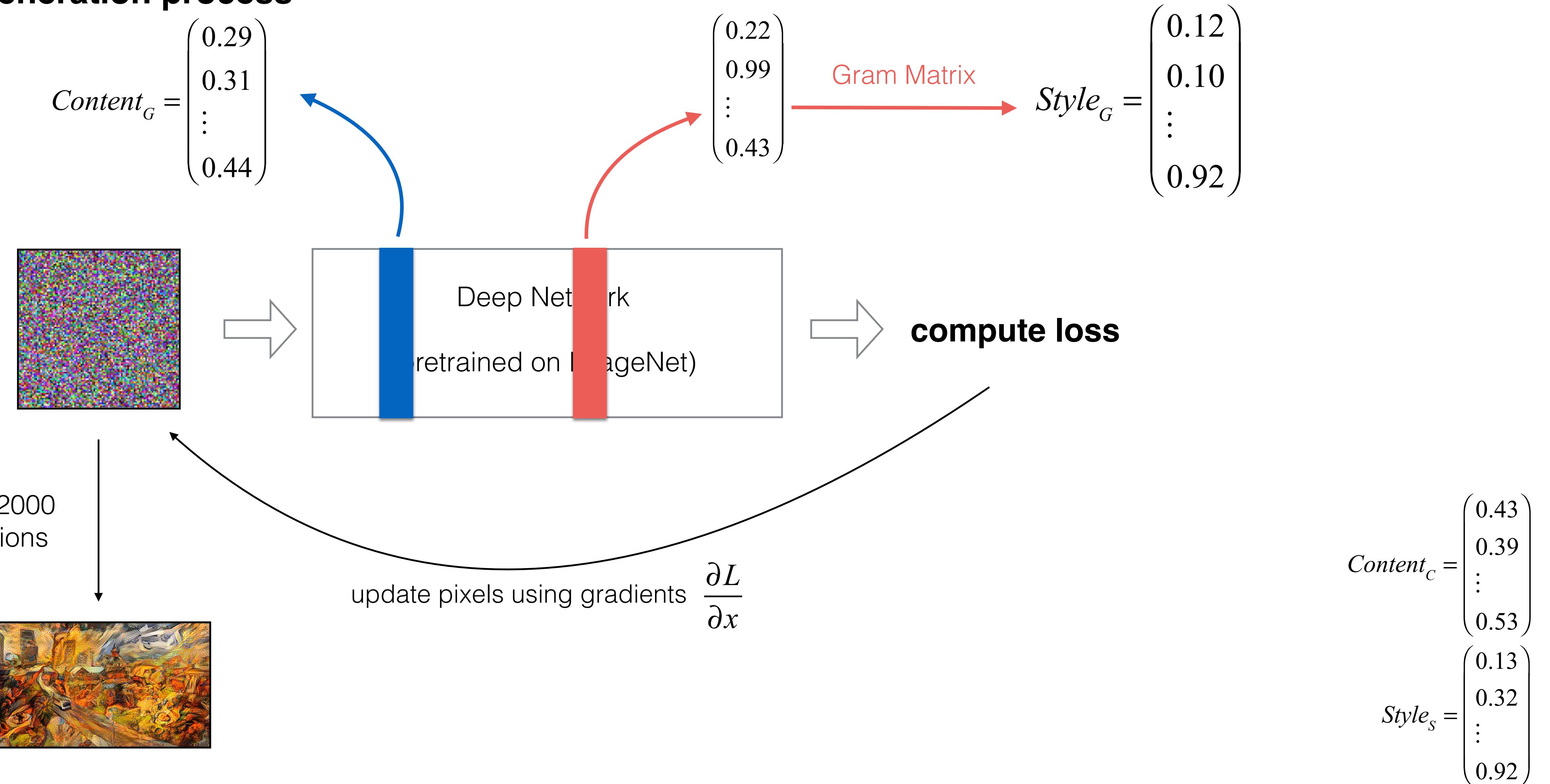
$$Content_C = \begin{pmatrix} 0.43 \\ 0.39 \\ \vdots \\ 0.53 \end{pmatrix}$$
$$Style_S = \begin{pmatrix} 0.13 \\ 0.32 \\ \vdots \\ 0.92 \end{pmatrix}$$

Gram Matrix



Art generation (Neural Style Transfer)

Image generation process



Art generation (Neural Style Transfer)

Which loss should we minimize?

$$L = \|Content_C - Content_G\|_2^2$$

$$- \|Style_S - Style_G\|_2^2$$

$$L = \|Style_S - Style_G\|_2^2$$

$$+ \|Content_C - Content_G\|_2^2$$

$$L = \|Style_S - Style_G\|_2^2$$

$$- \|Content_C - Content_G\|_2^2$$

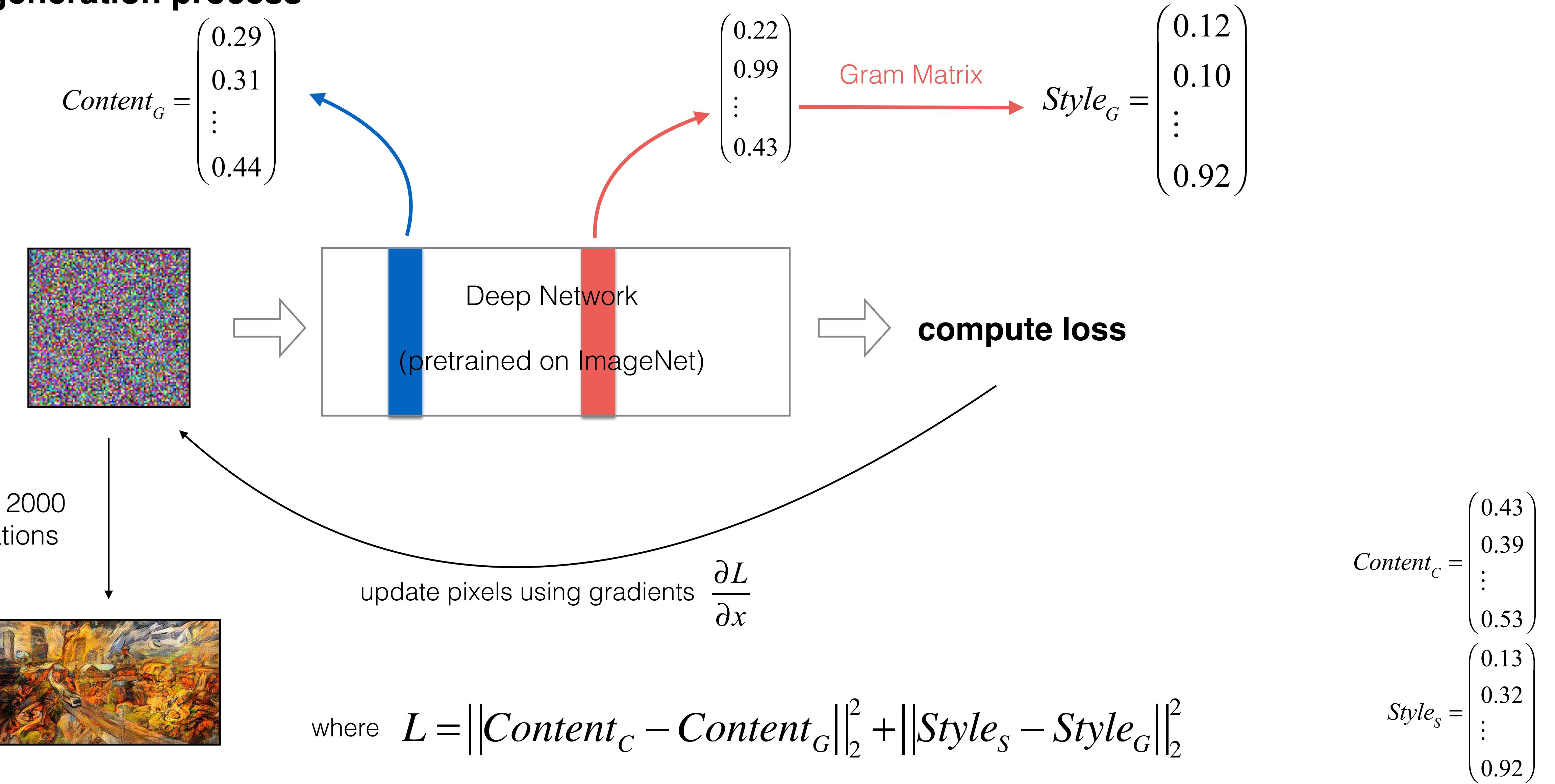
A

B

C

Art generation (Neural Style Transfer)

Image generation process





Content image



In the style of Hilma af Klint



In the style of Jamini Roy



In the style of Claude Monet



In the style of Yayoi Kusama



In the style of Eiichiro Oda



In the style of Salvador Dali



In the style of Piet Mondrian



In the style of Pablo Picasso

Trigger word detection

Goal: Given a 10sec audio speech, detect the word “activate”.

1. Data?

A bunch of 10s audio clips

Distribution?

2. Input?

$x = \text{A 10sec audio clip}$



Resolution? (sample rate)

3. Output?

$y = 0 \text{ or } y = 1$

Let's have an experiment!



$$y = 1$$



$$y = 0$$



$$y = 1$$



Trigger word detection

Goal: Given a 10sec audio speech, detect the word “activate”.

1. Data?

A bunch of 10s audio clips

Distribution?

2. Input?

$x = \text{A 10sec audio clip}$



Resolution? (sample rate)

3. Output?

$y = 0 \text{ or } y = 1$

Last Activation?

$y = 00..0000\mathbf{1}00000..000$

sigmoid

$y = 00..0000\mathbf{1}..1000..000$

(sequential)

4. Architecture ?

Sounds like it should be a RNN

5. Loss?

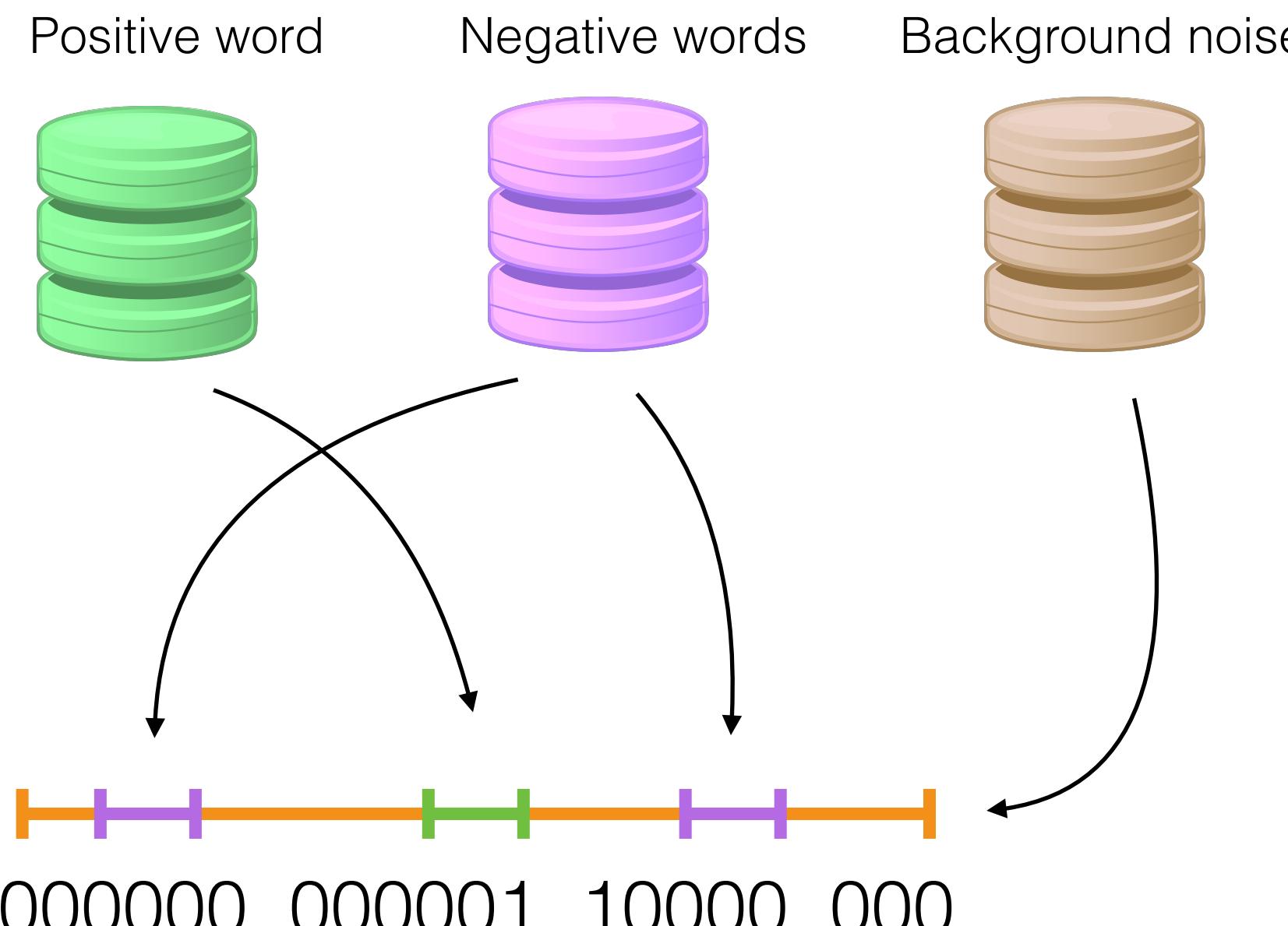
$$L = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}))$$

(sequential)

Trigger word detection

What is critical to the success of this project?

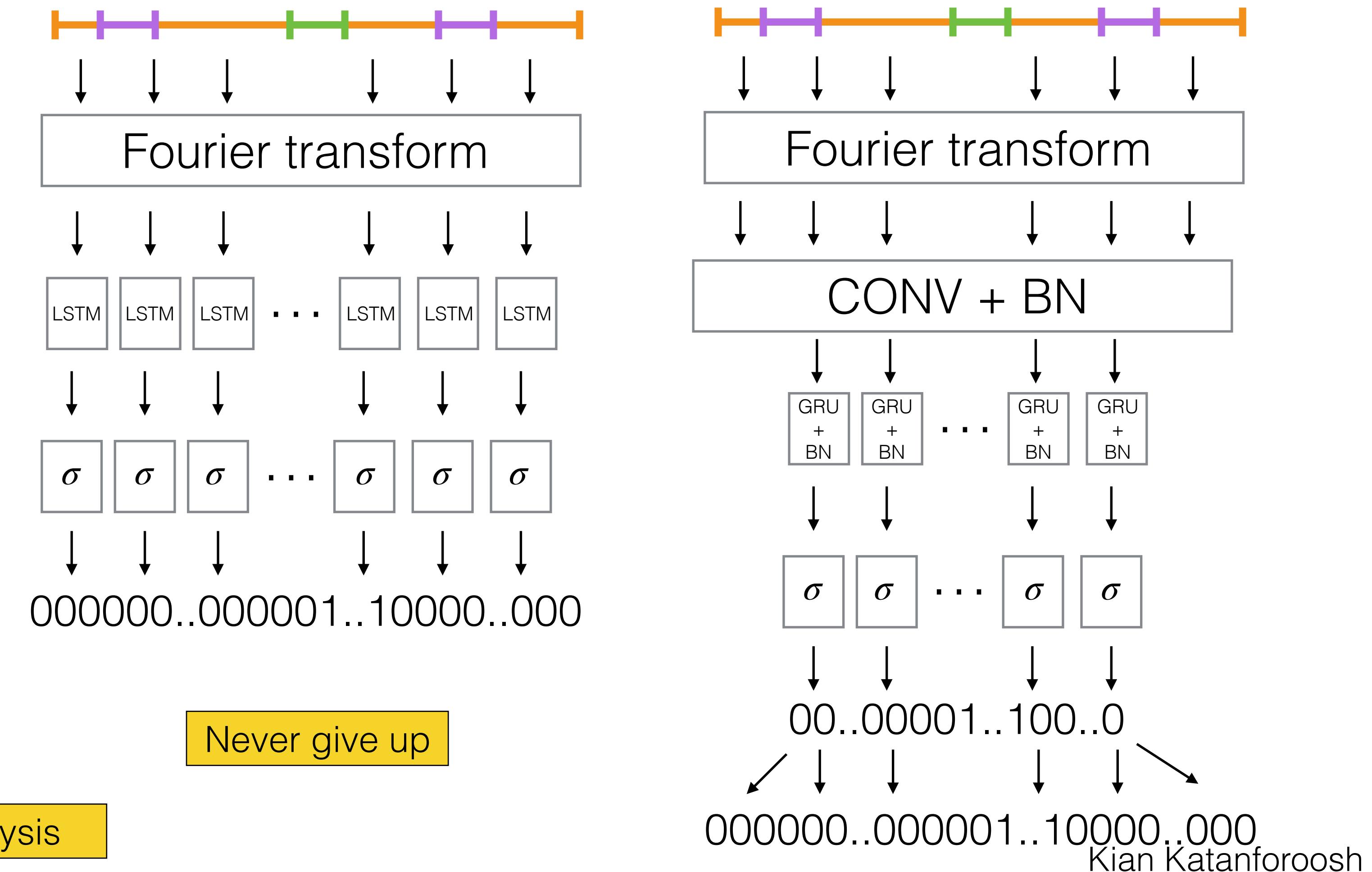
1. Strategic data collection/ labelling process



Automated labelling

+ Error analysis

2. Architecture search & Hyperparameter tuning



Never give up

Kian Katanforoosh

Featured in the Magazine “the Most Beautiful Loss functions of 2015”

$$\begin{aligned}
& \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
& + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
& + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2 \\
& + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2 \\
& + \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
\end{aligned}$$

Duties for next week

For Tuesday 04/23, 8am:

C2M1

- Quiz: Practical aspects of deep learning
- Programming Assignment: Initialization
- Programming Assignment: Regularization
- Programming Assignment: Gradient Checking

C2M2

- Quiz: Optimization Algorithms
- Programming Assignment: Optimization

Others:

- TA project mentorship (mandatory this week)
- Friday TA section (04/19)
- Fill-in AWS Form to get GPU credits for your projects