Applied Text Analytics &
Natural Language Processing

with Dr. Mahdi Roozbahani
& Wafa Louhichi

Deep Learning
Convolutional Neural Network (CNN) – Part 1

Some of the slides are based on Ming Li (University of Waterloo – Deep Learning Part) with some modifications

GT

Learning Objectives

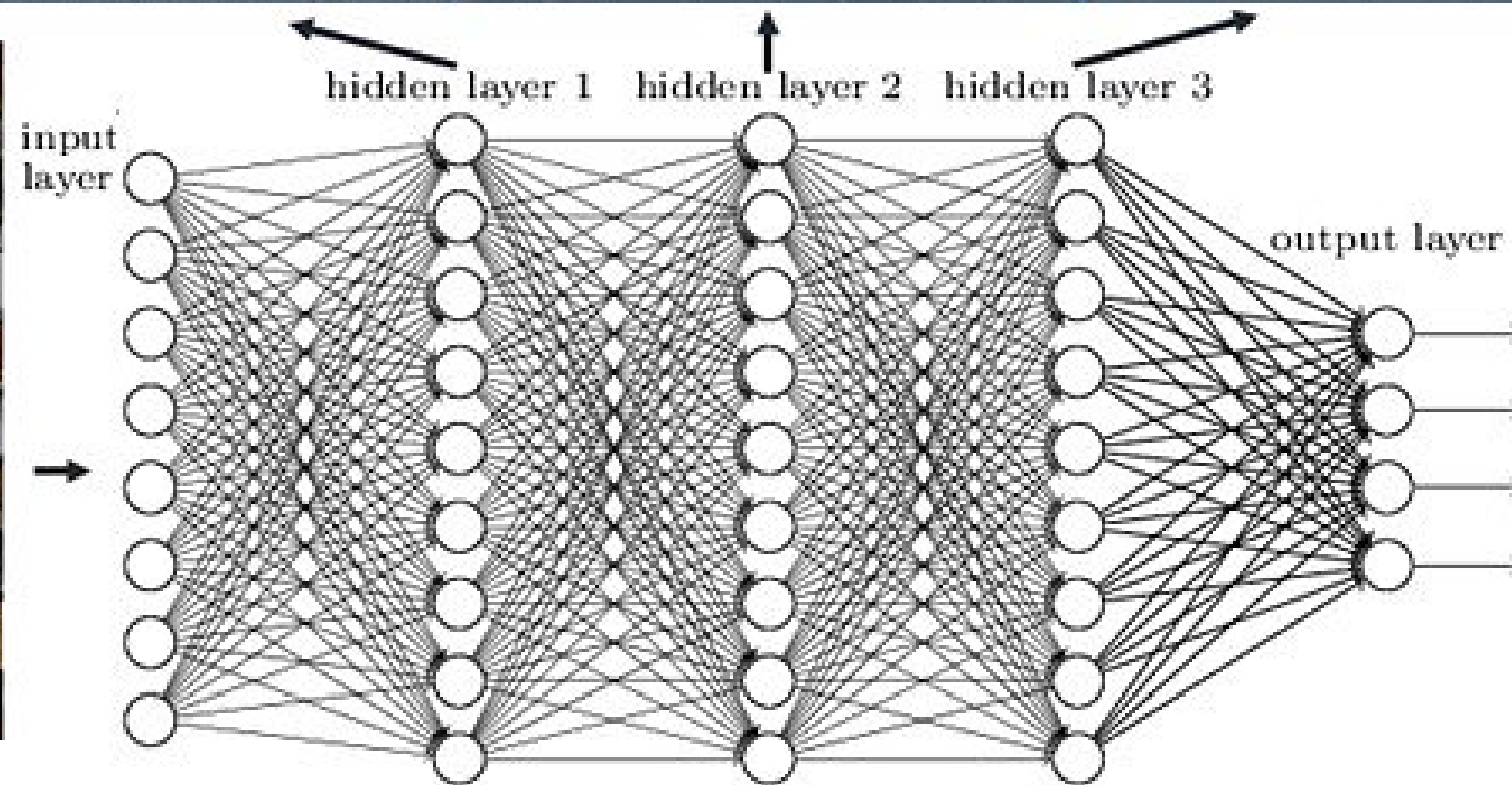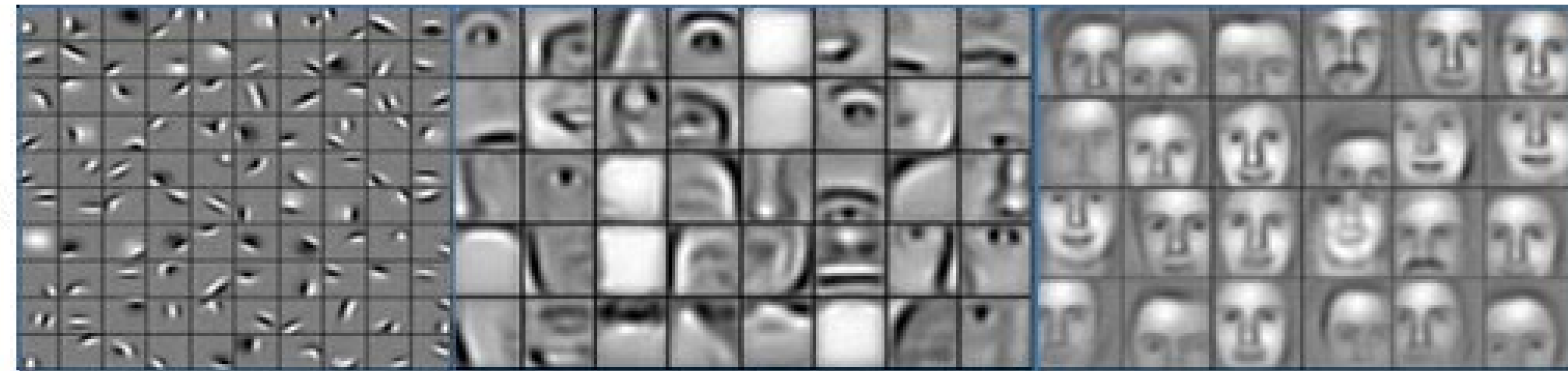In this lesson, you will learn a deep learning model called CNN

- Deep Learning
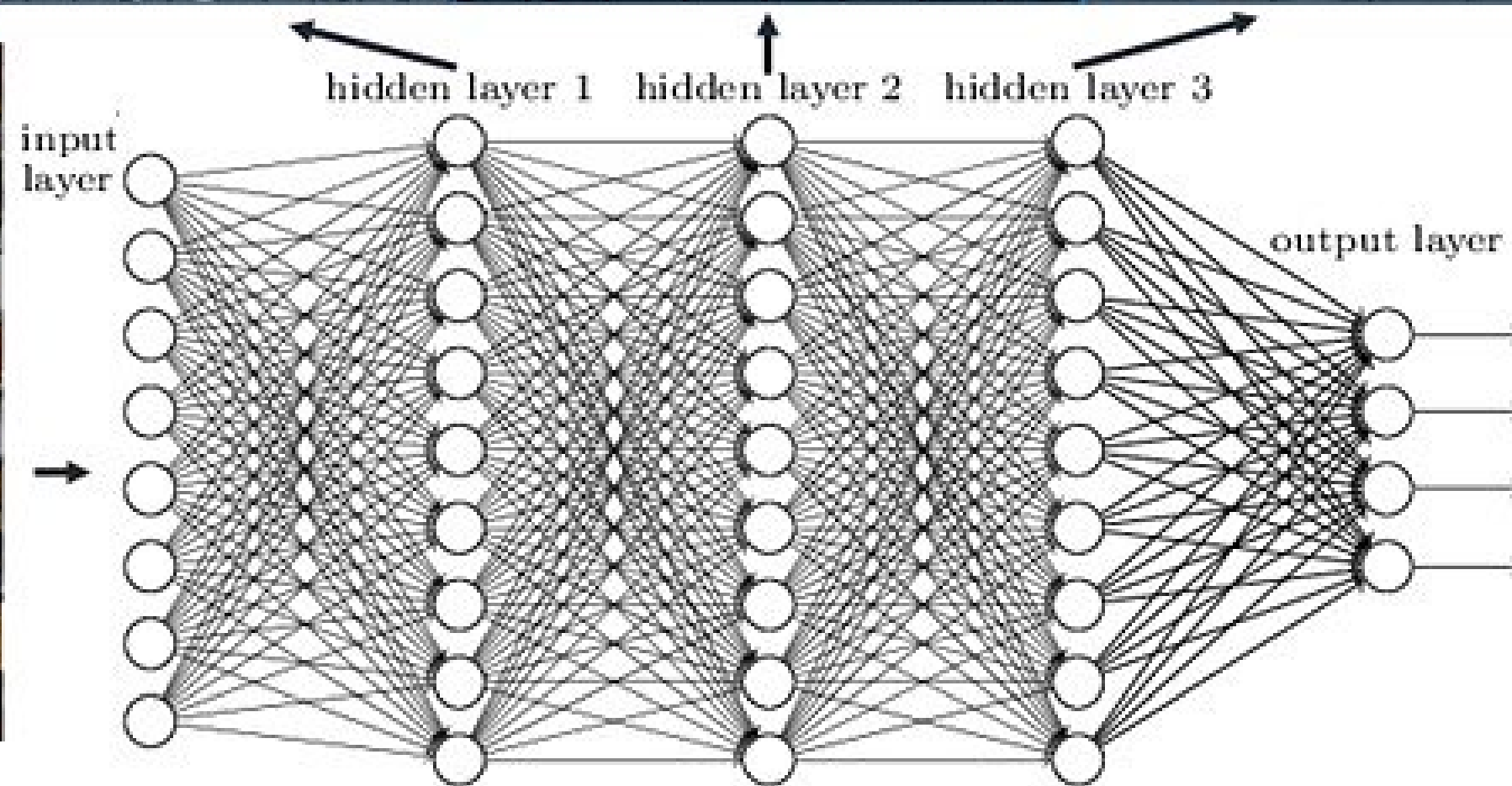
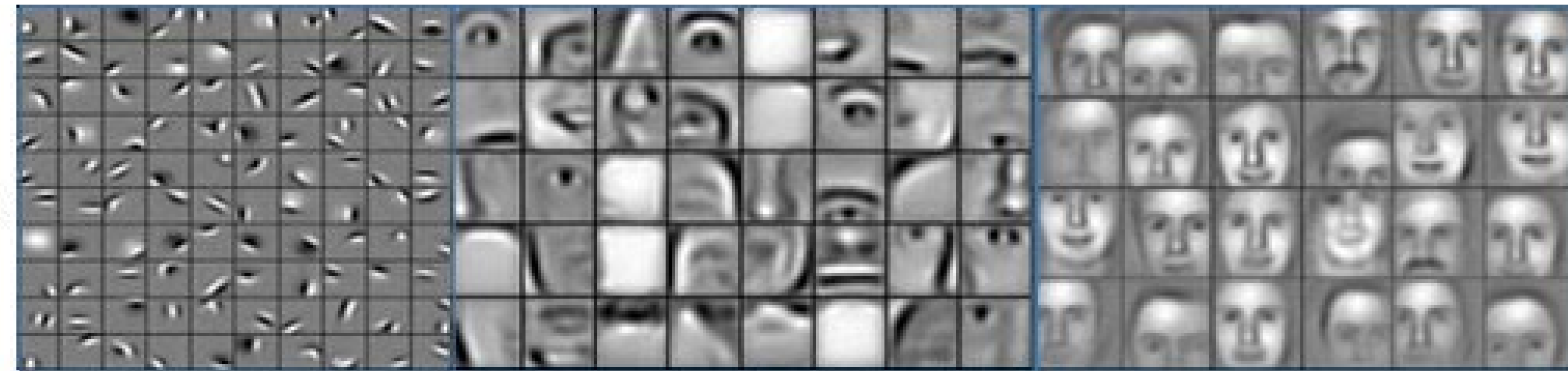- Convolutional Neural Network (CNN)

What is a Deep Learning Model? What is the Main Issue When a Model Becomes Deep?

https://medium.com/@fenjiro/face-id-deep-learning-for-face-recognition-324b50d916d1

What is a Deep Learning Model? What is the Main Issue When a Model Becomes Deep?

Smaller Network: ANN

- We know it is good to learn a small model
- From this fully connected model, do we really need all the edges?
- Can some of these be shared?
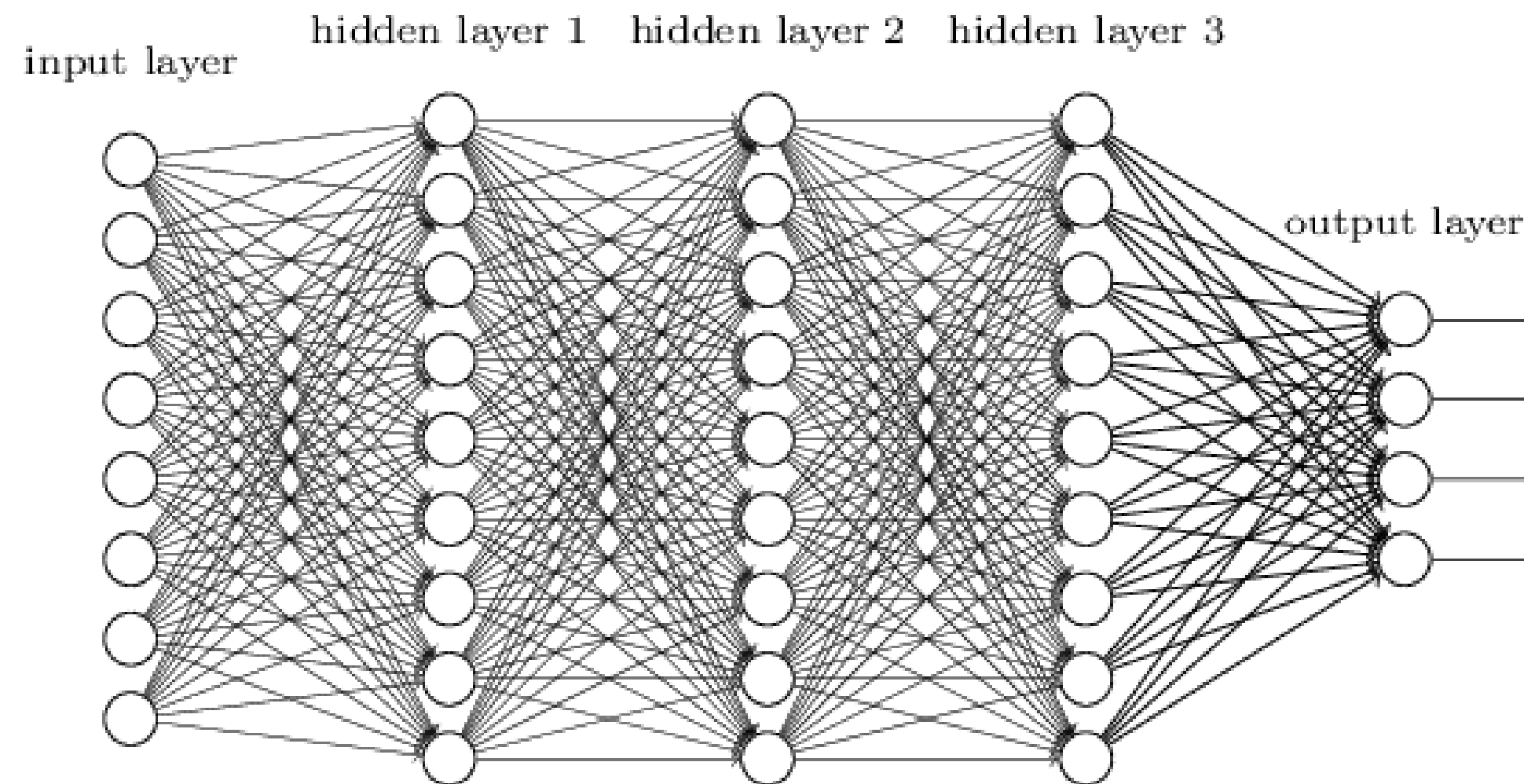
Consider Learning an Image:

- Some patterns are much smaller than the whole image

Can represent a small region with fewer parameters

"beak" detector

Consider Learning an Image:
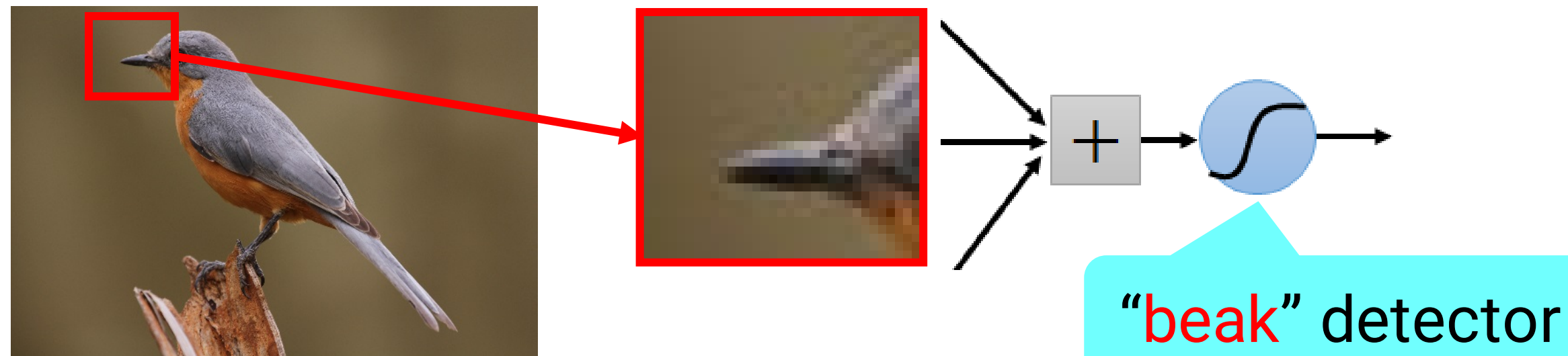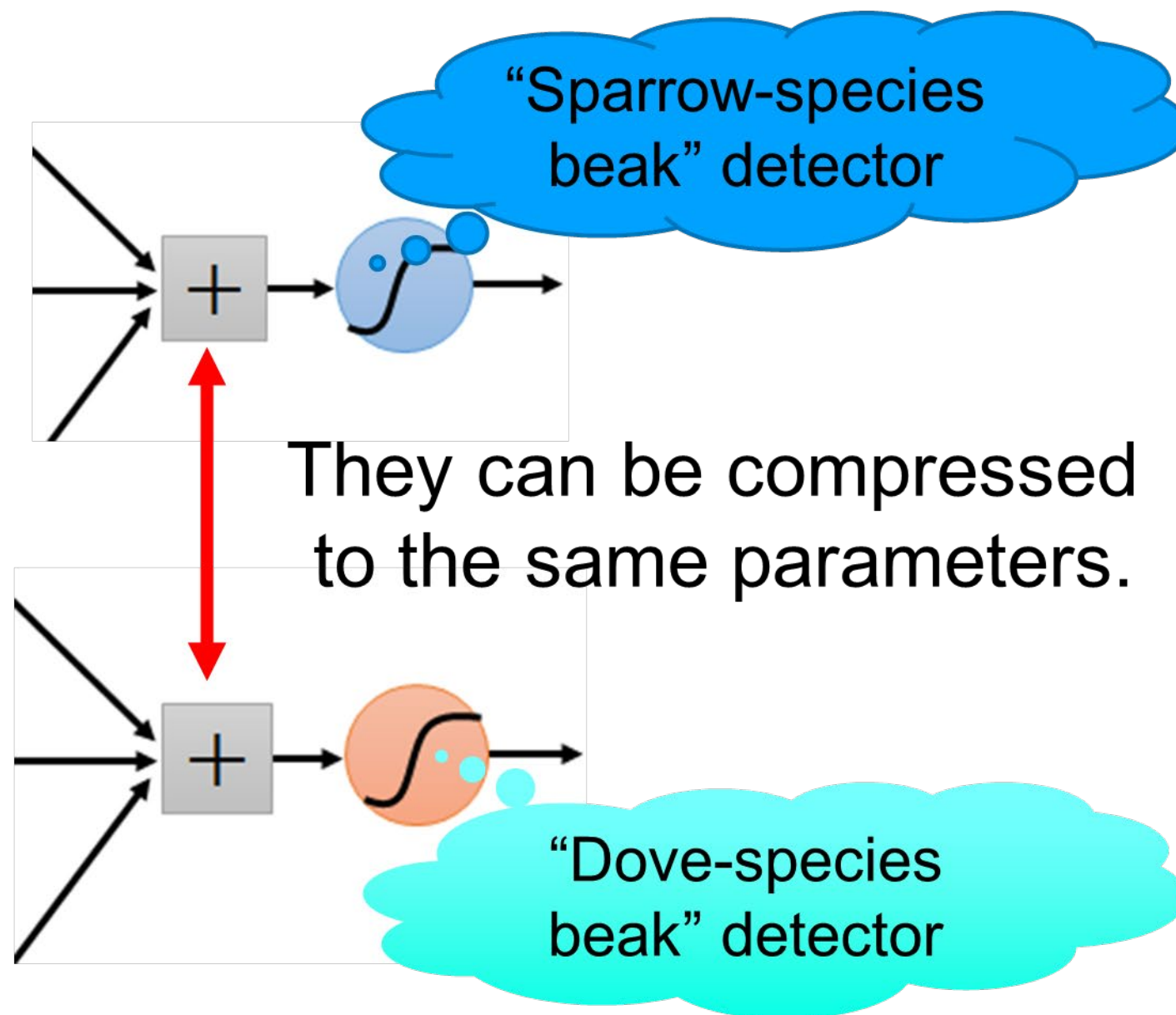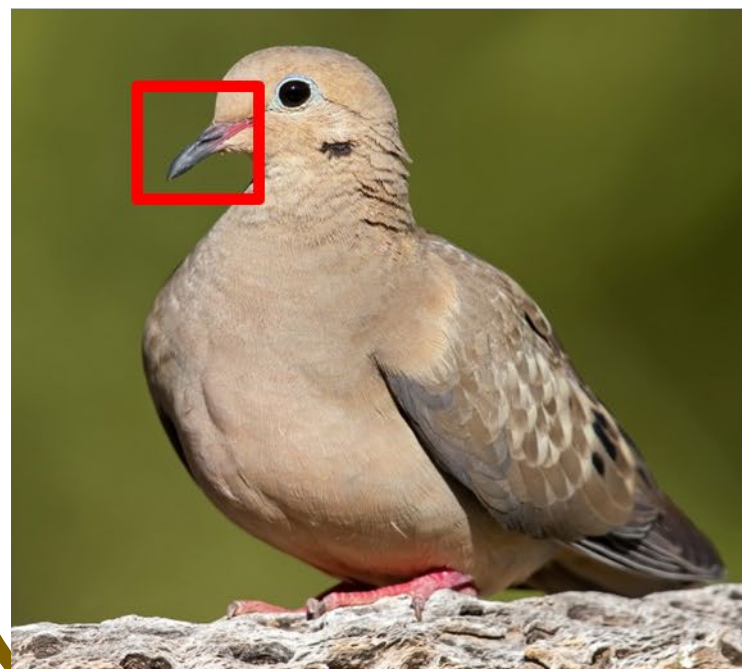
- Some patterns are much smaller than the whole image

Can represent a small region with fewer parameters

"beak" detector
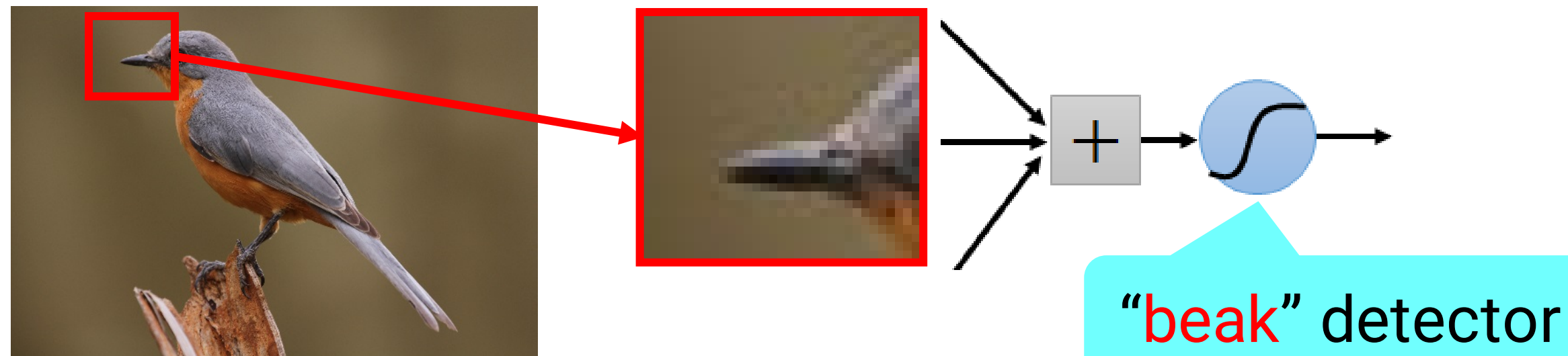
Why CNN Works Well for Images?
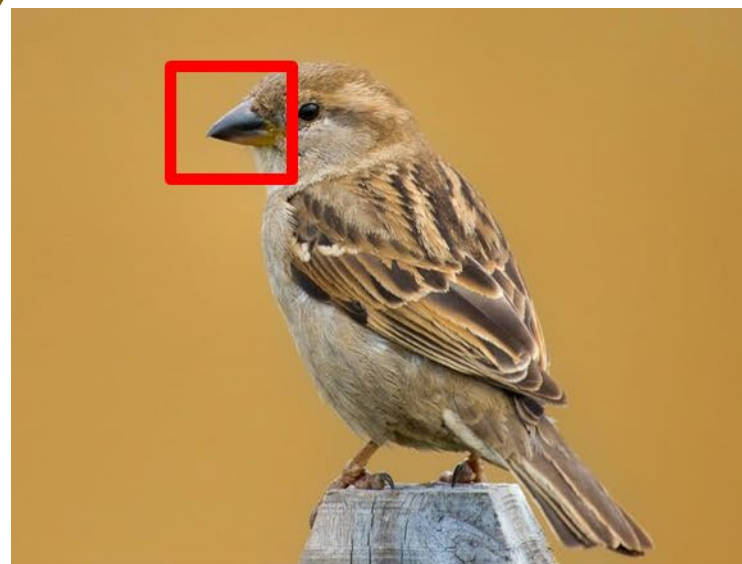
What about training a lot of such "small" detectors and each detector must "move around"?

Can we use CNN for NLP?

Yes, but we do need to prepare a dataset which is readable to a CNN model like an image which is a matrix

How to Prepare the Dataset for an NLP Problem?

For simplicity; let's say we have 3 documents (3 sample data points). These documents are collected from Emails:

- **Wafa and Mahdi teach NLP class**
- **NLP is neat**
- **CNN is a good model**

- Our vocabulary vector has 12 unique or distinct words; i.e. wafa, and, mahdi, …
- The longest document has 6 words (Note that CNN needs all the datapoints to have the same size)

Appling One-Hot Encoding Data Pre-Processing Technique

wafa   = [1,0,0,0,0,0,0,0,0,0,0,0]
and   = [0,1,0,0,0,0,0,0,0,0,0,0]
mahdi = [0,0,1,0,0,0,0,0,0,0,0,0]
⋮

- Each unique word will be a vector with a length equal to 12
- There are 12 unique words in our vocabulary vector

Converting Documents into a Matrix

Let's convert the first document:

- **Wafa and Mahdi teach NLP class**

wafa	1	0	0	0	0	0	0	0	0	0	0	0
and	0	1	0	0	0	0	0	0	0	0	0	0
mahdi	0	0	1	0	0	0	0	0	0	0	0	0
teach	0	0	0	1	0	0	0	0	0	0	0	0
nlp	0	0	0	0	1	0	0	0	0	0	0	0
class	0	0	0	0	0	1	0	0	0	0	0	0

Converting Documents into a Matrix

Let's convert the second document:

- **NLP is neat**

nlp	0	0	0	0	1	0	0	0	0	0	0	0
is	0	0	0	0	0	1	0	0	0	0	0	0
neat	0	0	0	0	0	0	1	0	0	0	0	0
Zero padding	0	0	0	0	0	0	0	0	0	0	0	0
Zero padding	0	0	0	0	0	0	0	0	0	0	0	0
Zero padding	0	0	0	0	0	0	0	0	0	0	0	0

- All documents will have the same size as the longest document in the corpus
- This is achieved by zero padding
- All documents can be converted in a same way