CNN

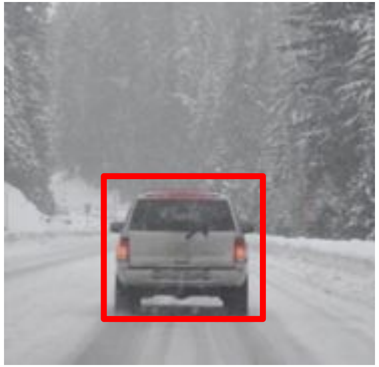
Week 3

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

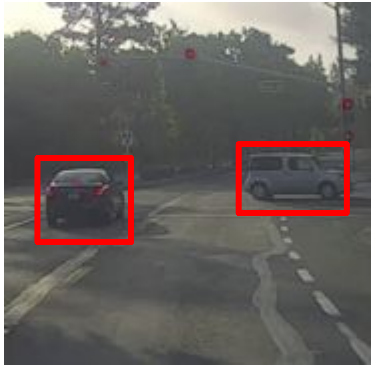
Classification, Localization and Detection

Image Classification

Classification with localization

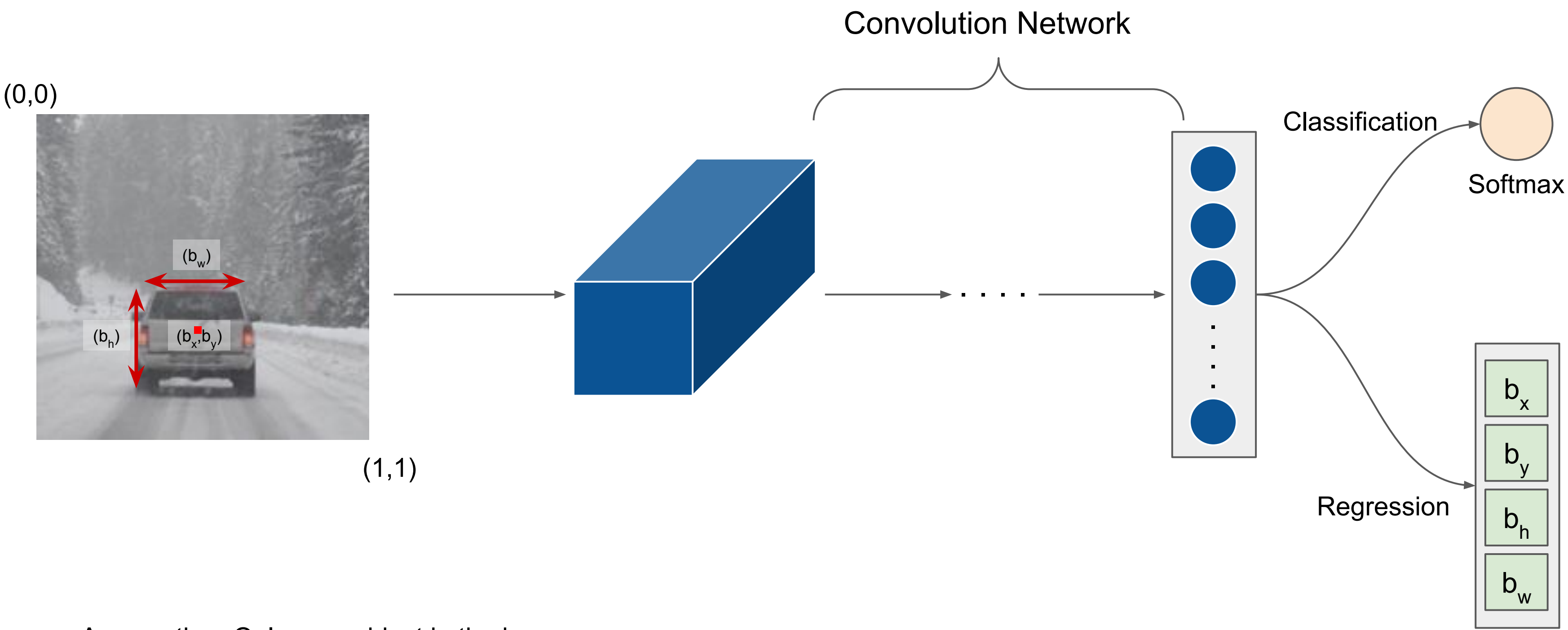


Detection



Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Classification with localization



Localization
Bounding box

Assumption: Only one object in the image

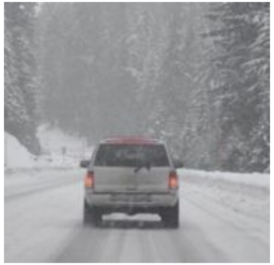
Define the output

3 Object Categories

Pedestrian

Car

Motorcycle



Probability of Object being present

Logistic Regression/Binary Entropy

Bounding box values

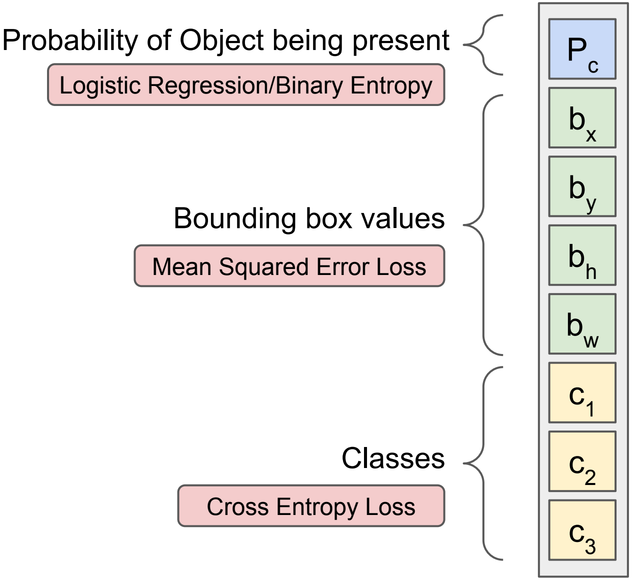
Mean Squared Error Loss

Classes

Cross Entropy Loss

P_c	1	0
b_x	0.5	NA
b_y	0.7	NA
b_h	0.3	NA
b_w	0.4	NA
c_1	0	NA
c_2	1	NA
c_3	0	NA

Define the output



Probability of Object being present

Logistic Regression/Binary Entropy

Bounding box values

Mean Squared Error Loss

Classes

Cross Entropy Loss

If P_c != 0 *[Object is present in image]*

$$L(y`, y) = (P_c` - P_c)^2 + (b_x` - b_x)^2 + (b_y` - b_y)^2 + (b_h` - b_h)^2 + (b_w` - b_w)^2 + (c_1` - c_1)^2 + (c_2` - c_2)^2 + (c_3` - c_3)^2$$

If P_c = 0 *[Object is not present in image]*

$$L(y`, y) = (P_c` - P_c)^2$$

3 Object Categories

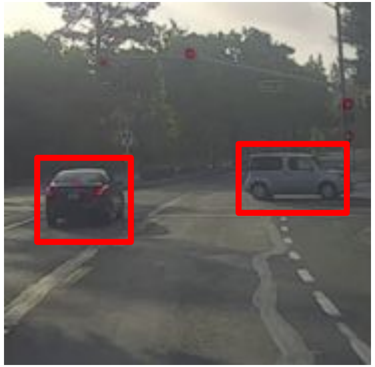
Pedestrian

Car

Motorcycle

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Object Detection



Training Data Set

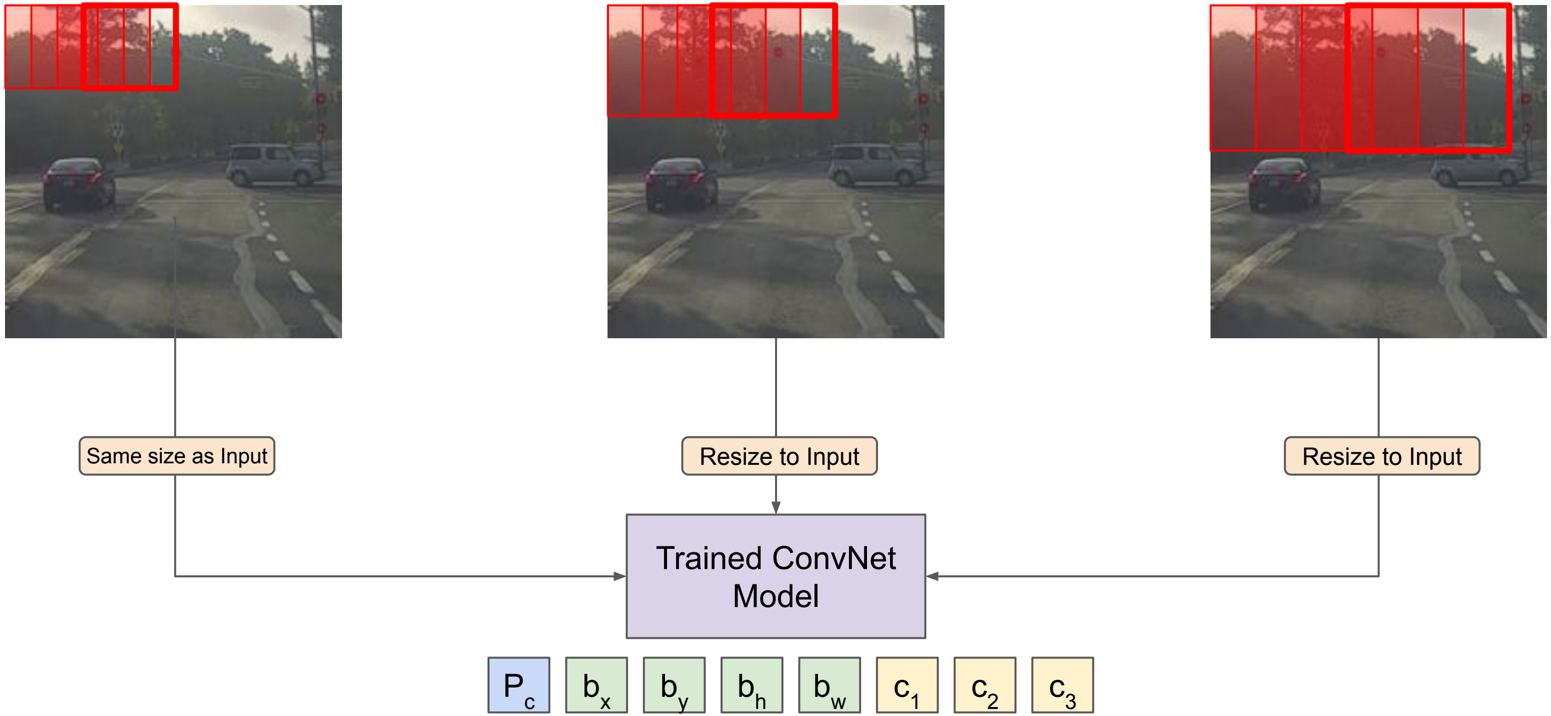
x	y
	1
	1
	1
	0
	0

Train a Convolutional Network using Closely-Cropped images.

Trained ConvNet Model

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Sliding Window Detection



Same size as Input

Resize to Input

Resize to Input

Trained ConvNet Model

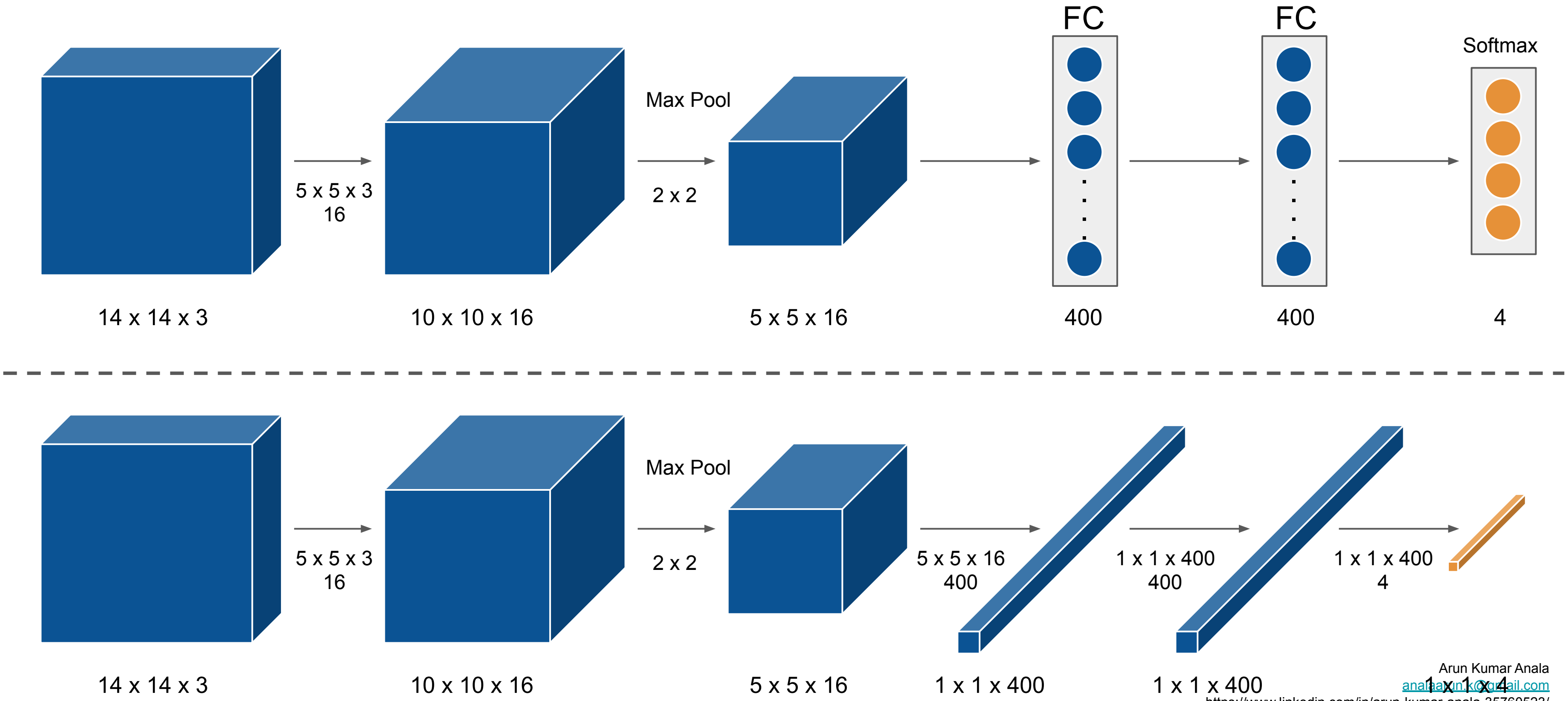
P_c	b_x	b_y	b_h	b_w	c_1	c_2	c_3

High Computation:
Sequential Processing of cropped images.
Lower Strides
Low dimensions of cropped images

Low Performance:
Higher Strides
Inaccurate boundary box values
Actual boundary box may not be square

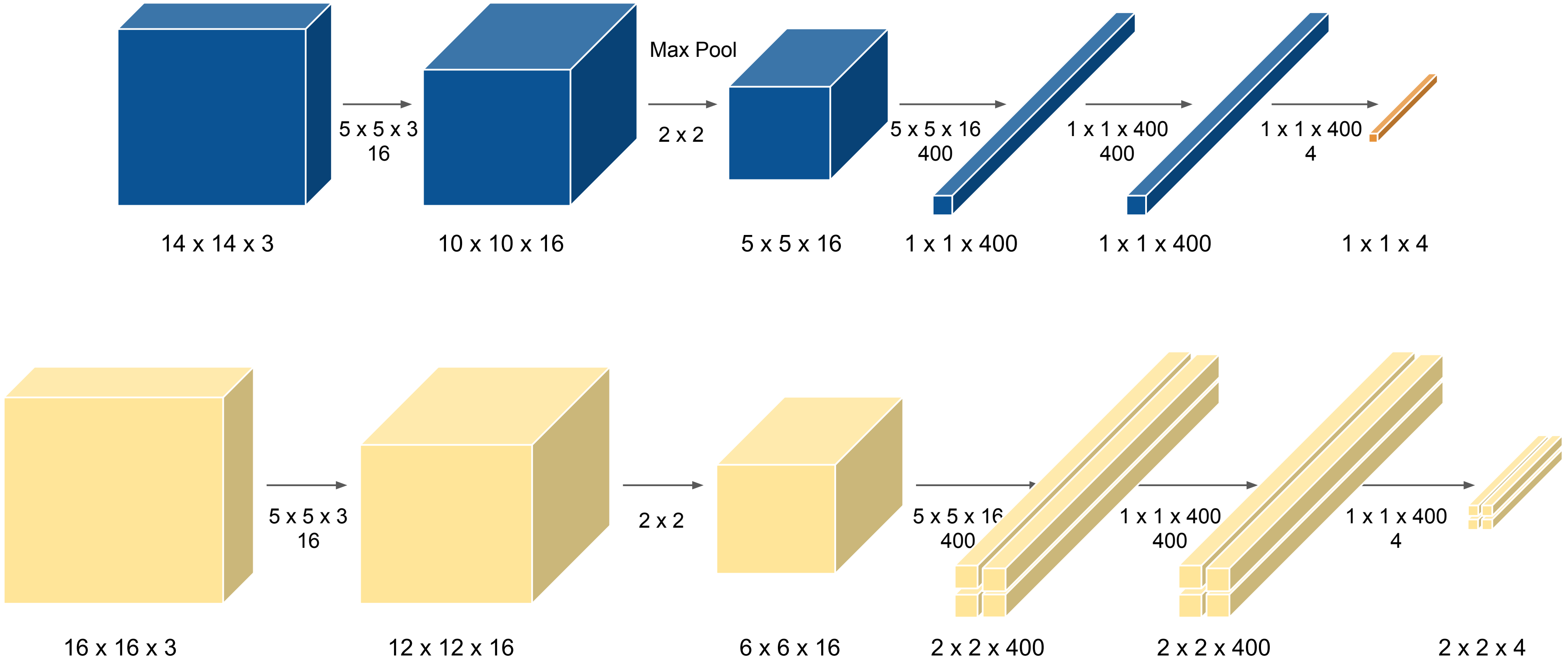
Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Sliding Window: Replace FC layers with Convolutional layers



Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Convolution Implementation of Sliding Window



14 x 14 x 3

5 x 5 x 3
16

10 x 10 x 16

Max Pool
2 x 2

5 x 5 x 16

5 x 5 x 16
400

1 x 1 x 400

1 x 1 x 400
400

1 x 1 x 400

1 x 1 x 400
4

1 x 1 x 4

16 x 16 x 3

5 x 5 x 3
16

12 x 12 x 16

2 x 2

6 x 6 x 16

5 x 5 x 16
400

2 x 2 x 400

1 x 1 x 400
400

2 x 2 x 400

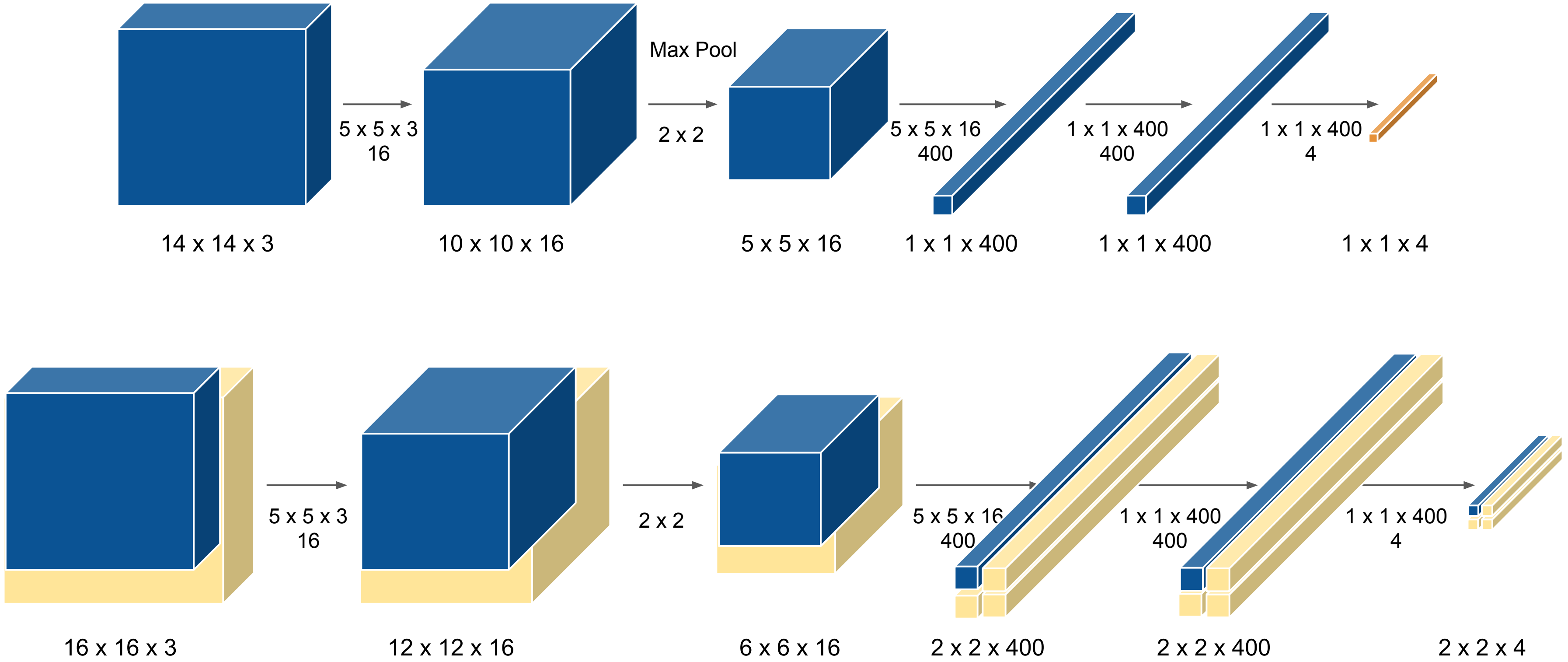
1 x 1 x 400
4

2 x 2 x 4

Input Image

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

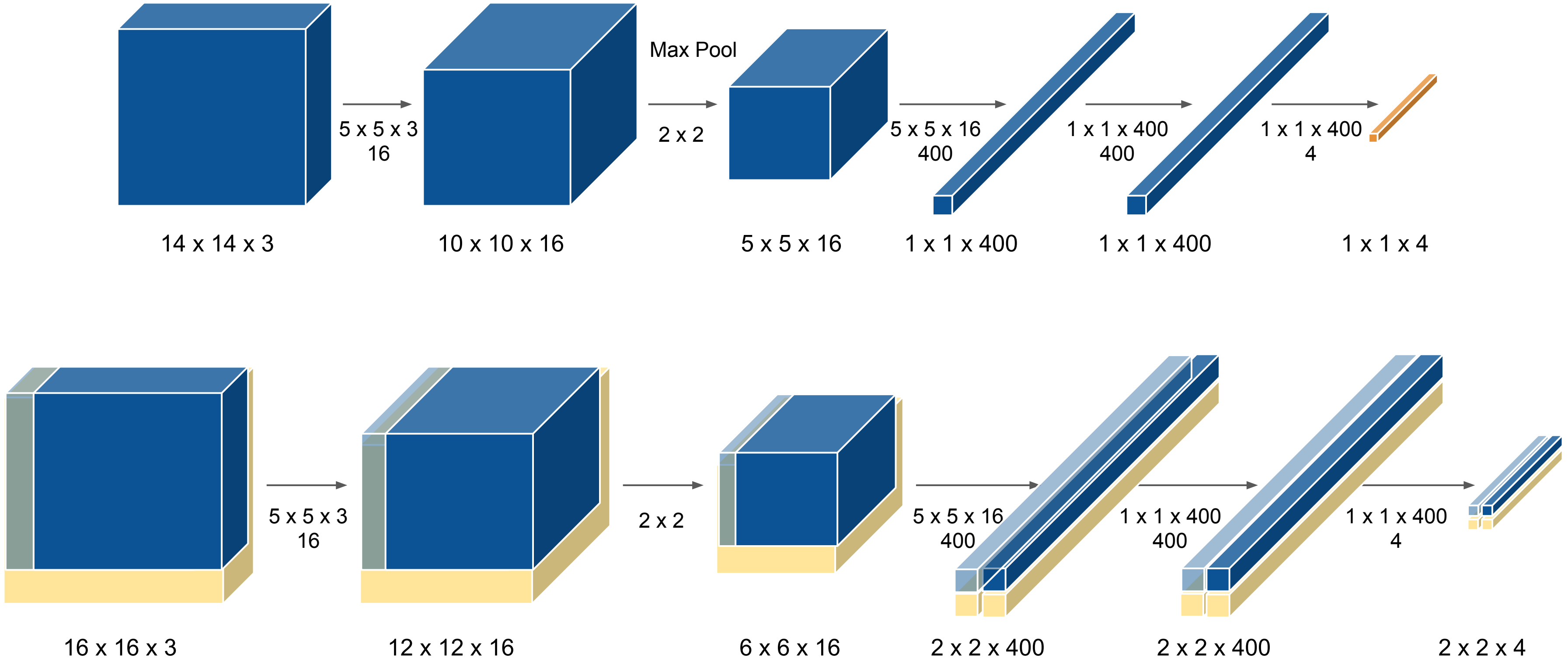
Convolution Implementation of Sliding Window



Input Image

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Convolution Implementation of Sliding Window



Max Pool

5 x 5 x 3
16

2 x 2

5 x 5 x 16
400

1 x 1 x 400
400

1 x 1 x 400
4

14 x 14 x 3

10 x 10 x 16

5 x 5 x 16

1 x 1 x 400

1 x 1 x 400

1 x 1 x 4

5 x 5 x 3
16

2 x 2

5 x 5 x 16
400

1 x 1 x 400
400

1 x 1 x 400
4

16 x 16 x 3

12 x 12 x 16

6 x 6 x 16

2 x 2 x 400

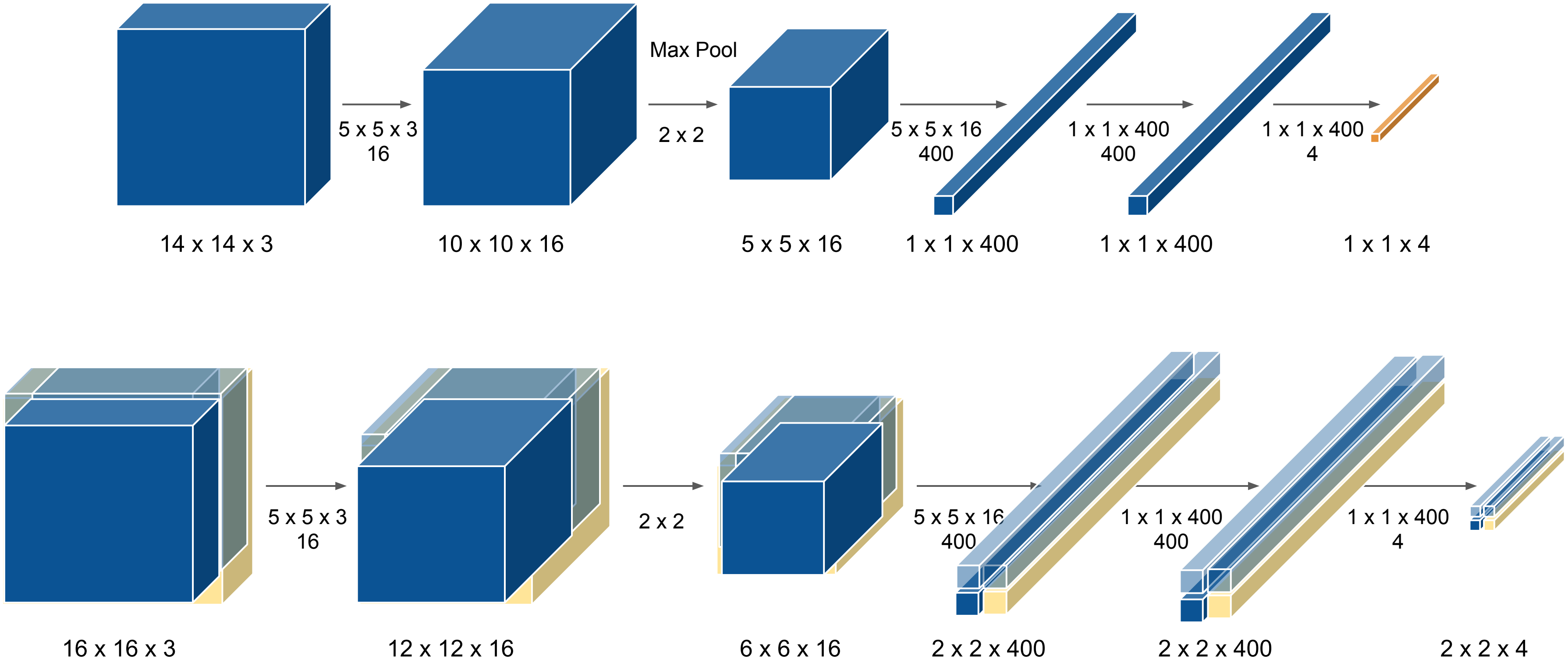
2 x 2 x 400

2 x 2 x 4

Input Image

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Convolution Implementation of Sliding Window



14 x 14 x 3

5 x 5 x 3
16

10 x 10 x 16

Max Pool
2 x 2

5 x 5 x 16

5 x 5 x 16
400

1 x 1 x 400

1 x 1 x 400
400

1 x 1 x 400

1 x 1 x 400
4

1 x 1 x 4

16 x 16 x 3

5 x 5 x 3
16

12 x 12 x 16

2 x 2

6 x 6 x 16

5 x 5 x 16
400

2 x 2 x 400

1 x 1 x 400
400

2 x 2 x 400

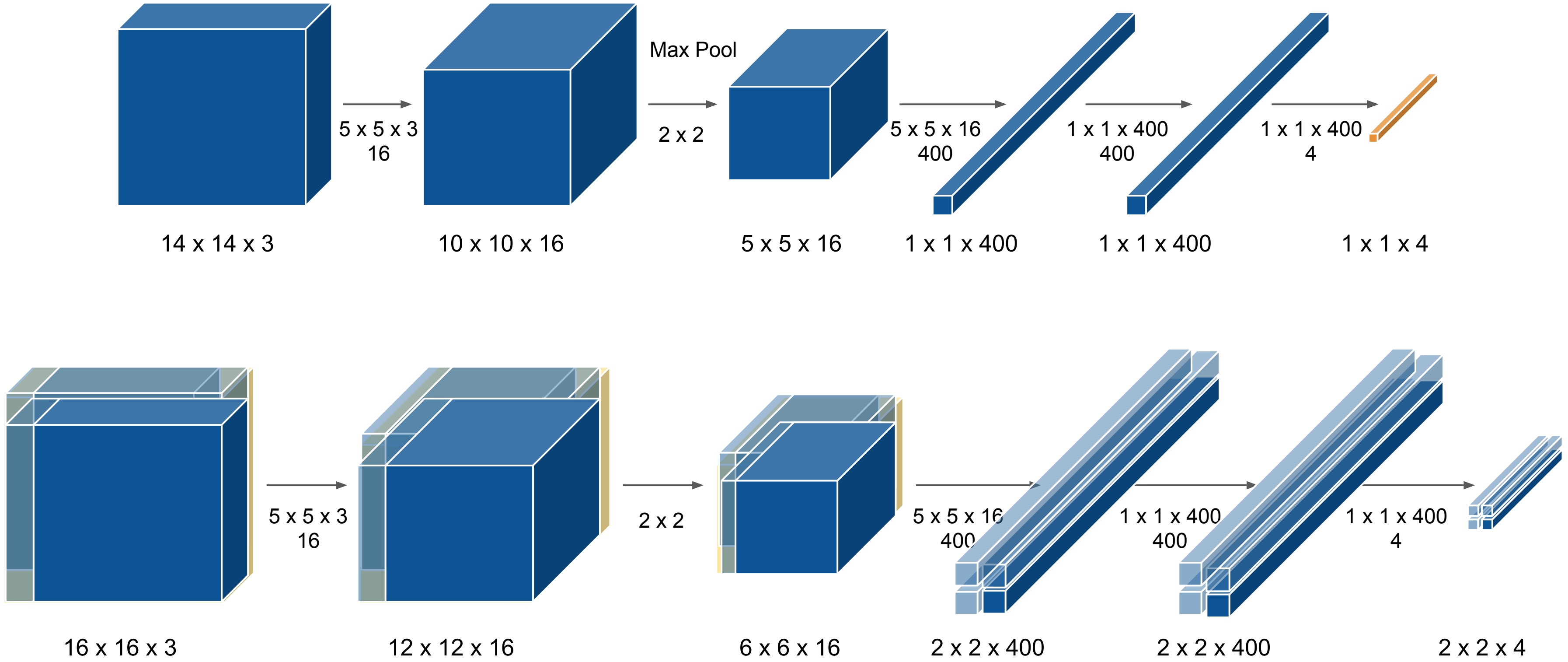
1 x 1 x 400
4

2 x 2 x 4

Input Image

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

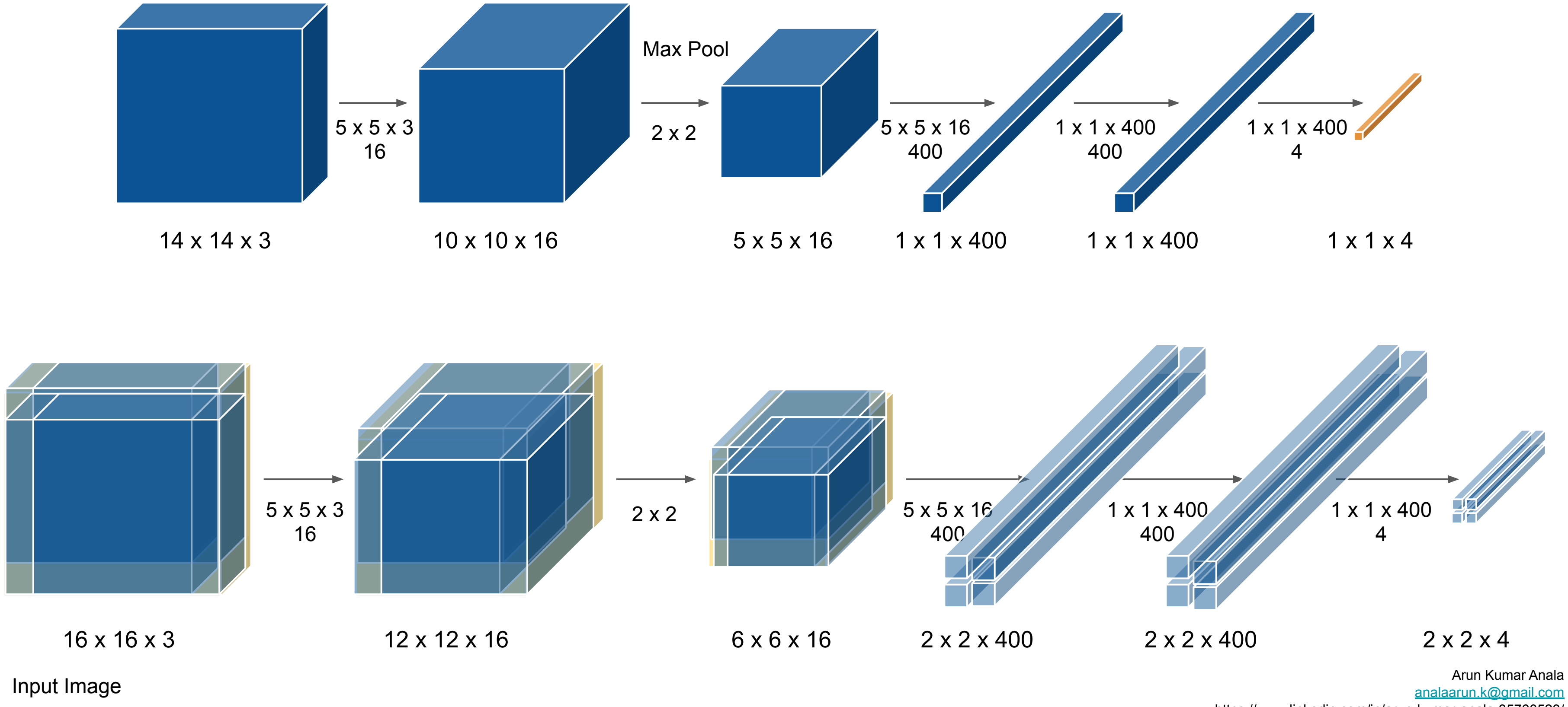
Convolution Implementation of Sliding Window



Input Image

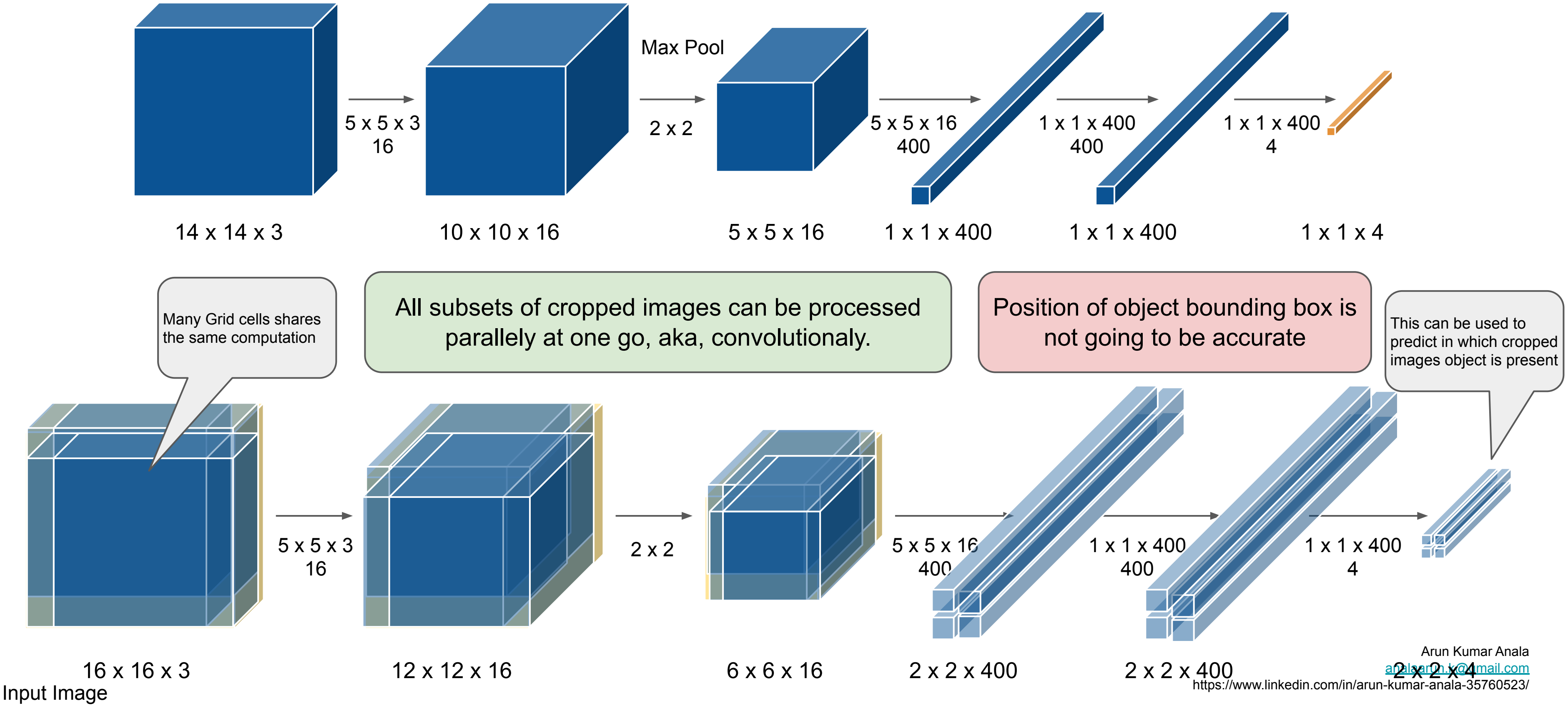
Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Convolution Implementation of Sliding Window



Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Convolution Implementation of Sliding Window



Many Grid cells shares the same computation

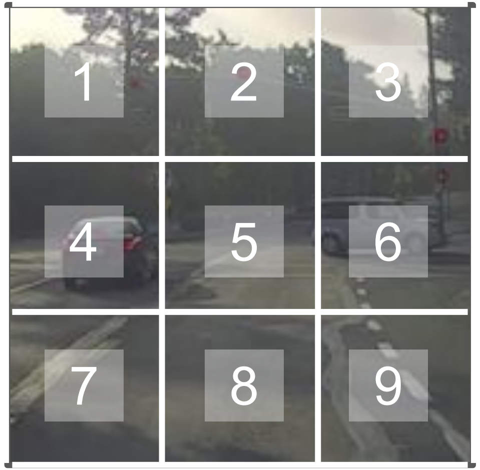
All subsets of cropped images can be processed parallely at one go, aka, convolutionaly.

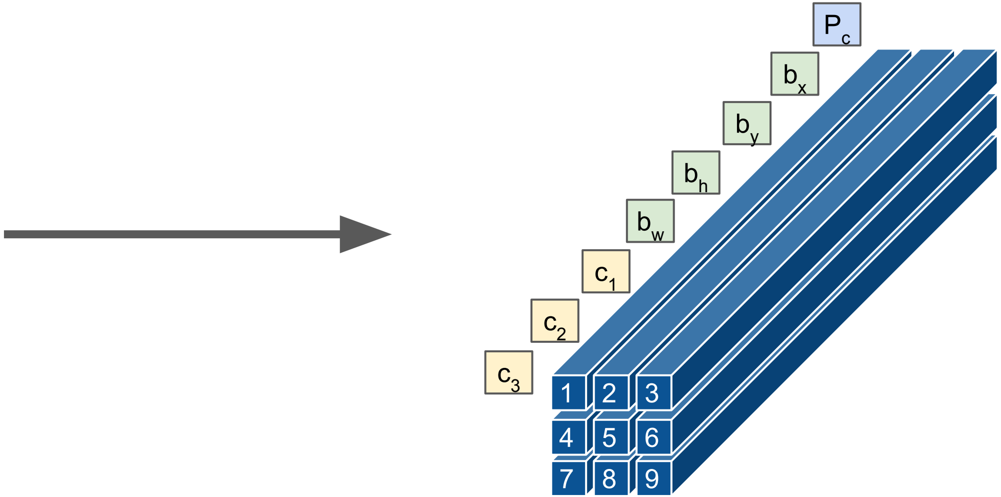
Position of object bounding box is not going to be accurate

This can be used to predict in which cropped images object is present

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

YOLO Algorithm (You Look Only Once)





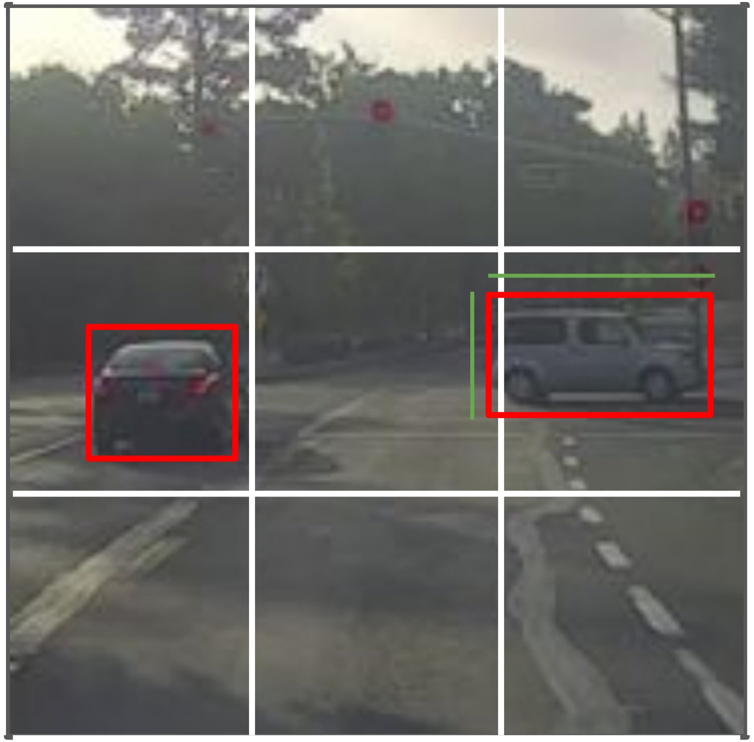
3 x 3 x 8

Output

Single Convolution Implementation
Real Time Object Detection.
# of Grids is a hyperparameter. Usually 19 x 19 considered as optimal to avoid image overlapping across grids
Assign the object to the grid that contains the center of the object

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

YOLO Algorithm (You Look Only Once)



Grid 6 Output

P_c	1
b_x	0.5
b_y	0.7
b_h	0.3
b_w	0.4
c_1	0
c_2	1
c_3	0

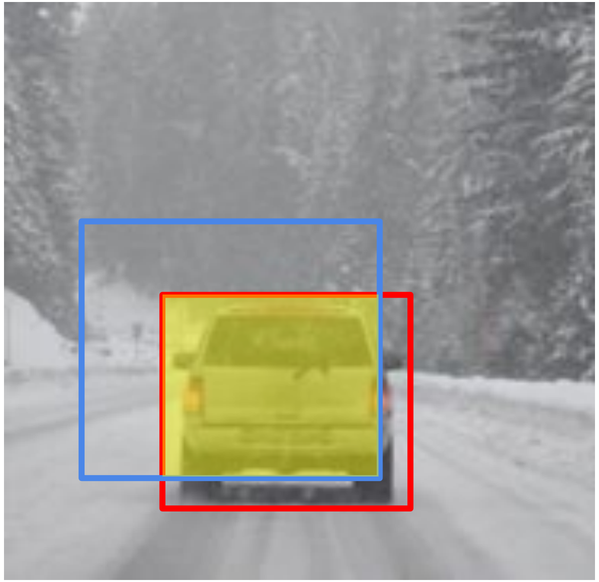
Output values are specified relative to the grid cell

b_x and b_y are always between 0 and 1

b_h and b_w could be > 1

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Measurement of Object Localization



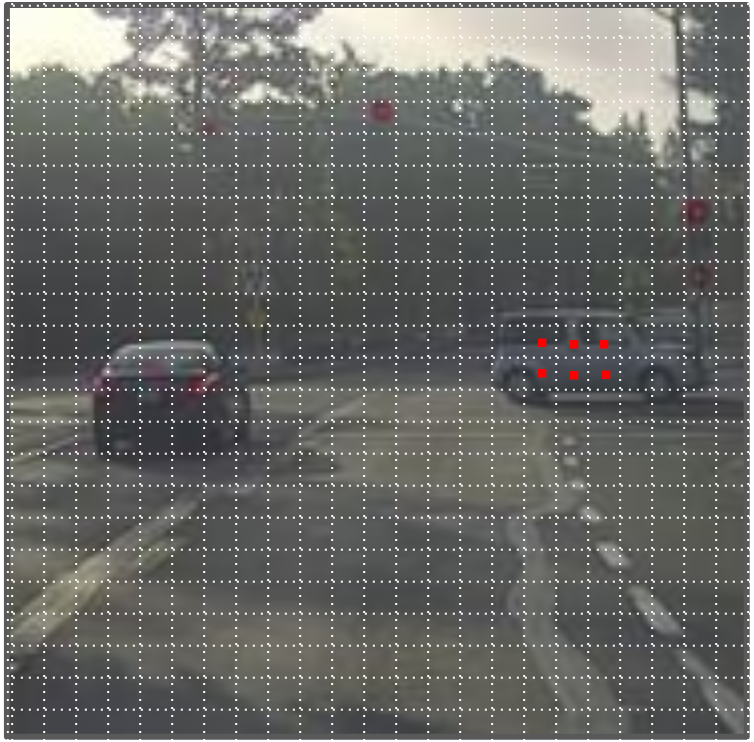
Intersection Over Union (IoU) = Size of 🟨 / Size of □

“Correct” if IoU >= 0.5

IoU is a measure of overlap between two bounding boxes

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Non-Max Suppression

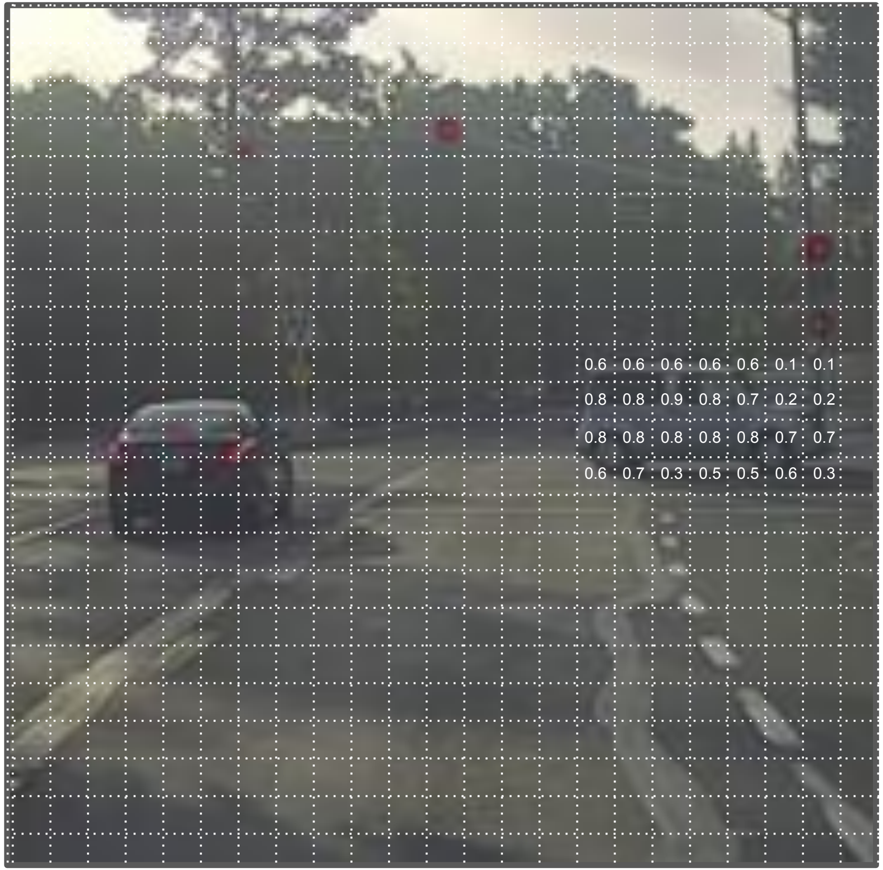


Multiple Grid Cells could detect the center of the car.

How to identify the number of cars in the image?

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Non-Max Suppression

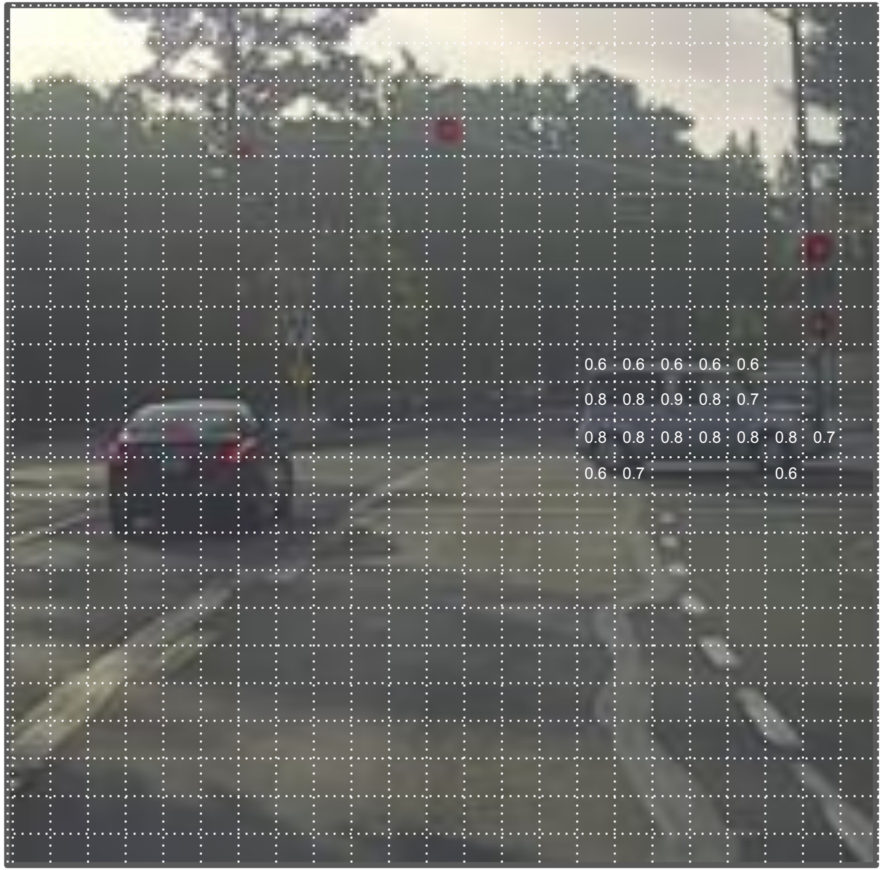


0.6	0.6	0.6	0.6	0.6	0.1	0.1
0.8	0.8	0.9	0.8	0.7	0.2	0.2
0.8	0.8	0.8	0.8	0.8	0.7	0.7
0.6	0.7	0.3	0.5	0.5	0.6	0.3

Each grid cell output is

P_c	b_x	b_y	b_h	b_w	c_1	c_2	c_3

Non-Max Suppression



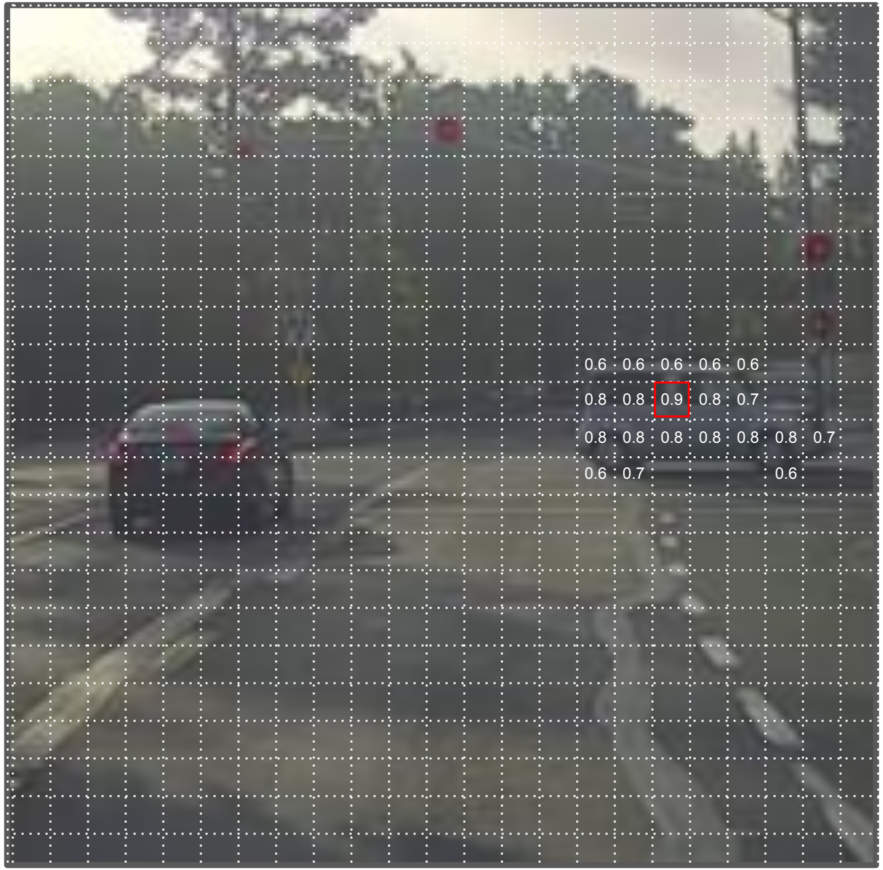
Each grid cell output is

P_c	b_x	b_y	b_h	b_w	c_1	c_2	c_3

Discard all boxes with P_c value of < 0.6.

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Non-Max Suppression



Each grid cell output is

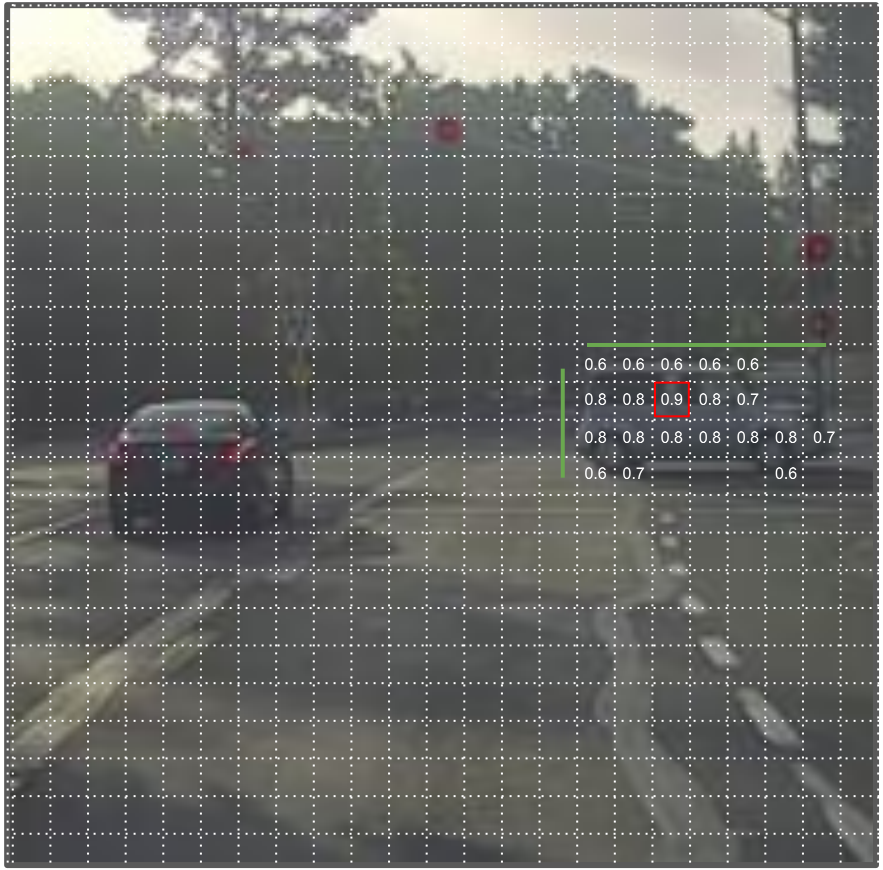
P_c b_x b_y b_h b_w c_1 c_2 c_3

Discard all boxes with P_c value of < 0.6.

```
While there are remaining boxes:
    Pick the box with the largest $P_c$ output as prediction
```

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Non-Max Suppression



Each grid cell output is

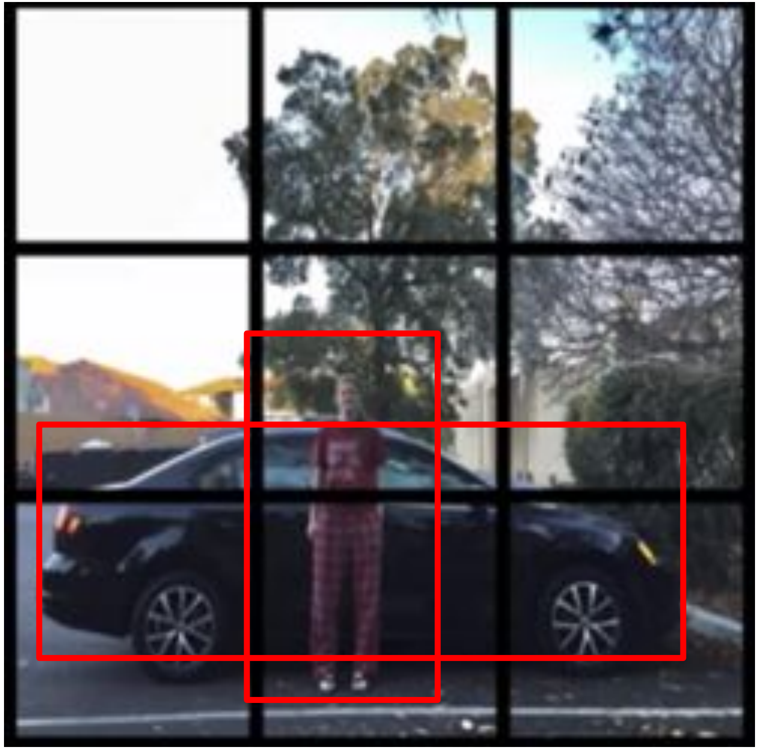
P_c	b_x	b_y	b_h	b_w	c_1	c_2	c_3

Discard all boxes with P_c value of < 0.6.

```
While there are remaining boxes:
    Pick the box with the largest $P_c$ output as prediction
    Discard any remaining boxes with IoU >= 0.5 with the box output in previous step.
```

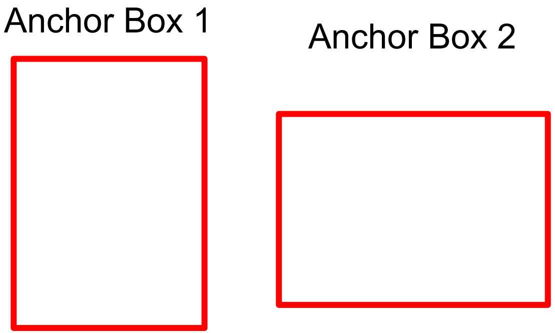
Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Anchor Boxes: Overlapping Objects

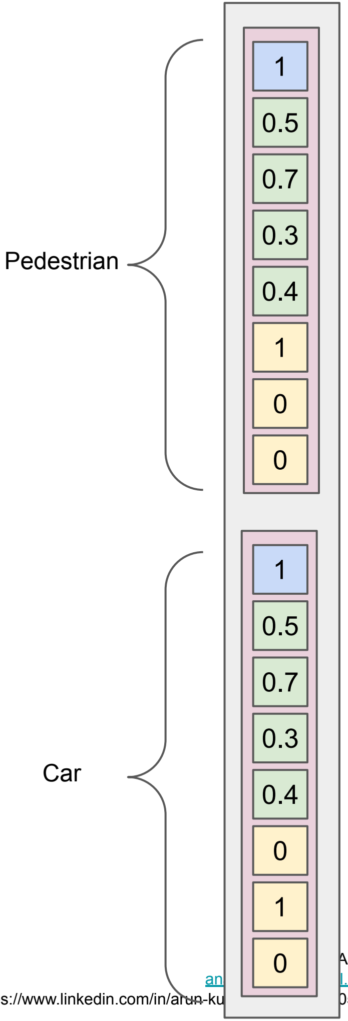


Use of Predefined Anchor boxes to detect multiple objects with in same grid cell.

Each grid cell would provide output for two anchor boxes

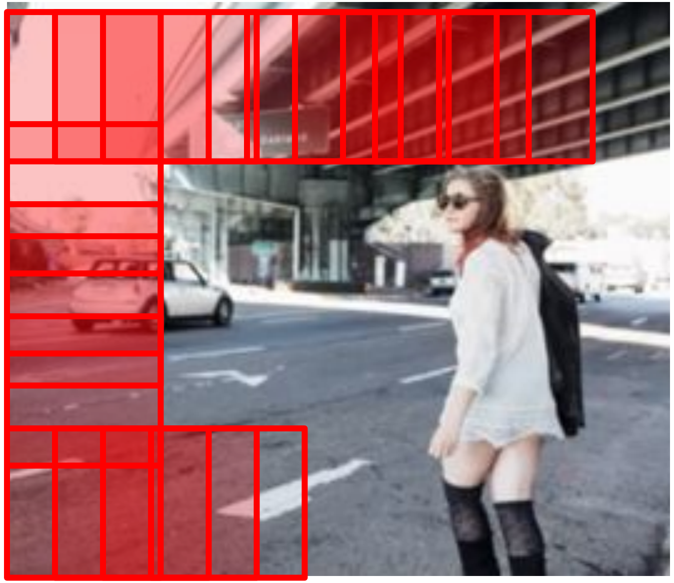


Output = 3 x 3 x 2 x 8



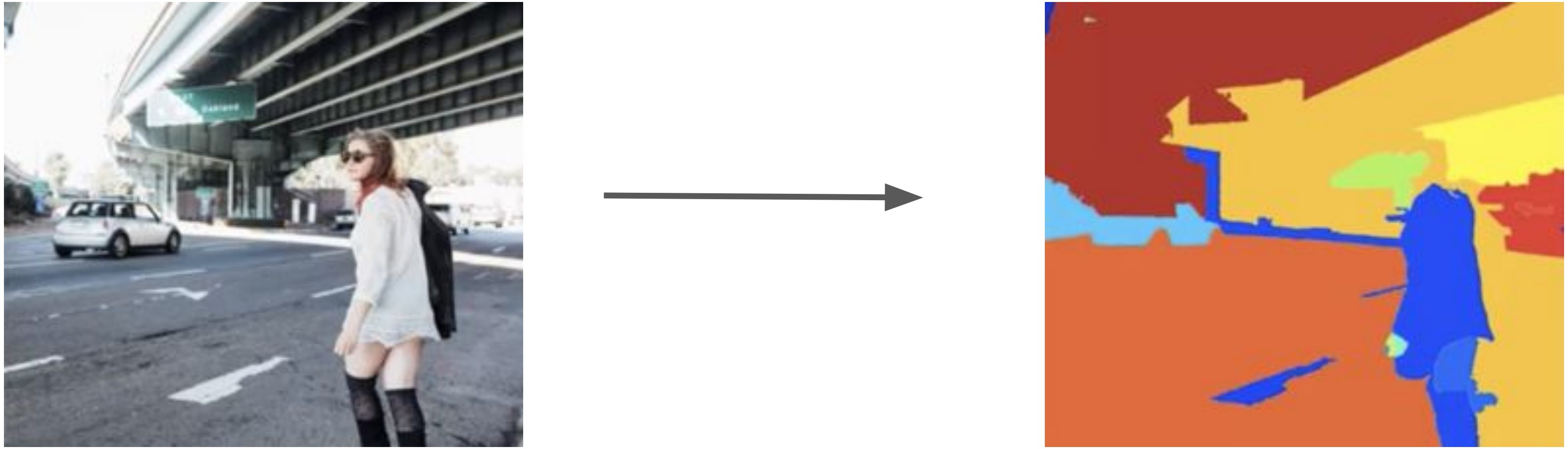
Anala
an l.com
https://www.linkedin.com/in/arun-ku 0523/

Region Proposal: R-CNN



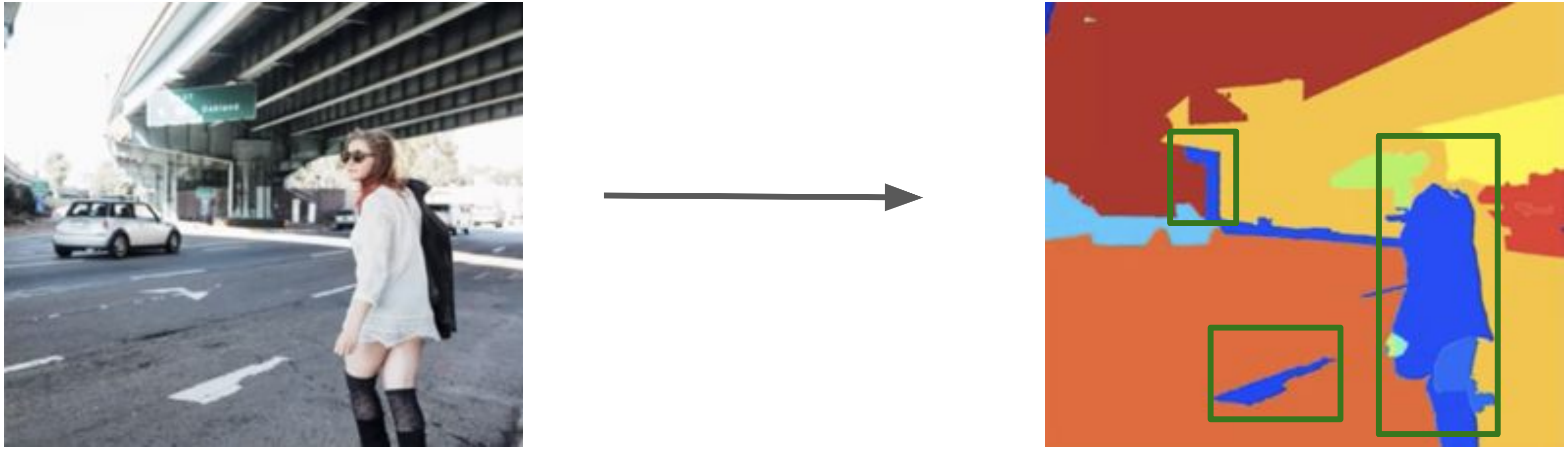
Disadvantage of Sliding window:
Classifies lot of cropped region that does not have any object.

Region Proposal: R-CNN



Uses Segmentation Algorithm to generate **blobs** in the image.

Region Proposal: R-CNN

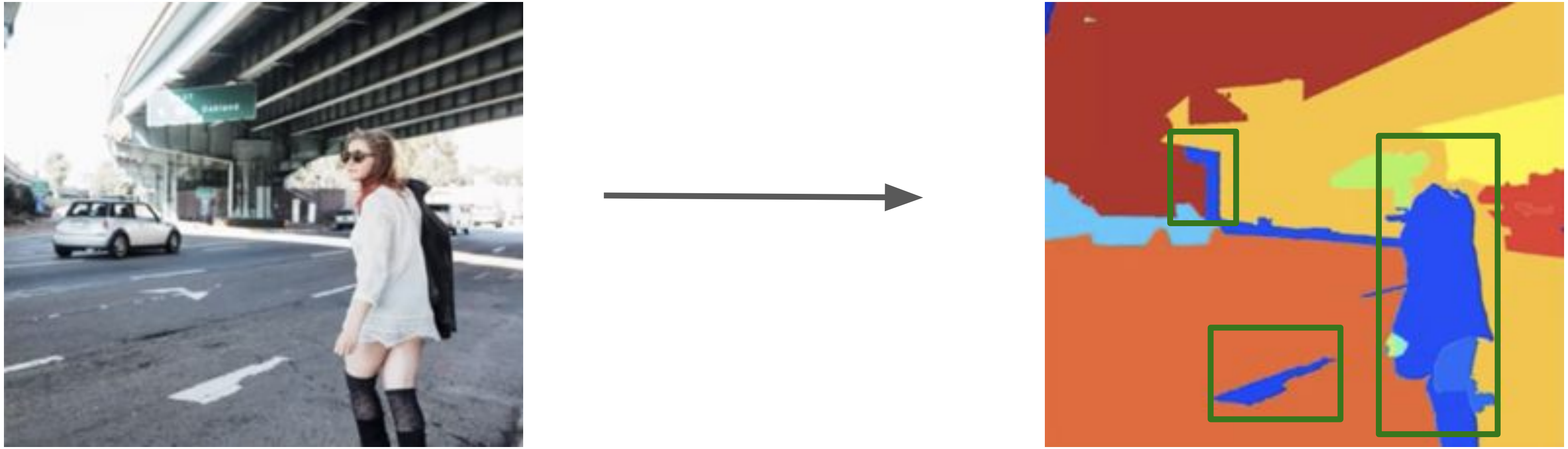


Uses Segmentation Algorithm to generate **blobs** in the images.

The bounding box of different scales is drawn across blobs, which is send for classification. If founds 200 bounding boxes, it send 2000 cropped regions for classification

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Region Proposal: R-CNN



Uses Segmentation Algorithm to generate **blobs** in the images.

The bounding box of different scales is drawn across blobs, which is send for classification. If founds 200 bounding boxes, it send 2000 cropped regions for classification

Disadvantage of R-CNN:
Segmentation algorithm is quite slow.

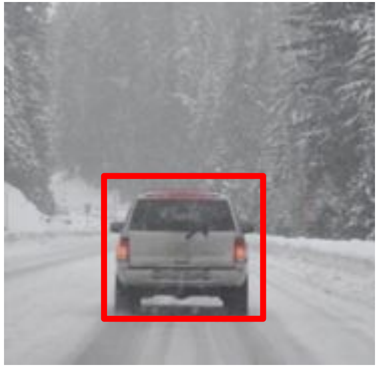
Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Region Proposal: Faster R-CNN

	⚡ R-CNN	⚡⚡⚡ FAST R-CNN	⚡⚡⚡⚡⚡⚡ FASTER R-CNN
Propose Region	Segmentation Algorithm to propose regions	Segmentation Algorithm to propose regions	⚡ Use Convolution Network to propose regions
Classification of Region	Sequential classification of proposed region	⚡ Convolution implementation to classify proposed regions	⚡ Convolution implementation to classify proposed regions

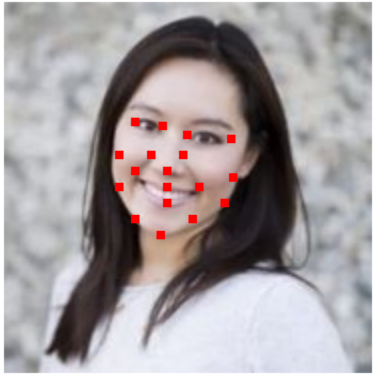
Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/

Landmark Detection



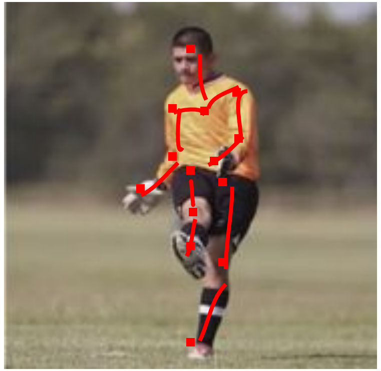
b_x b_y b_h b_w

Bounding Box



Recognize emotion

Used in AR (Augmented Reality)



Pose Detection

Each Coordinate is a Landmark

Output = 1 + # of Landmarks

1 for detecting object

Arun Kumar Anala
analaarun.k@gmail.com
https://www.linkedin.com/in/arun-kumar-anala-35760523/