NPTEL Week 12 Live Sessions

on Deep Learning (noc24_ee04)

A course offered by: Prof. Prabir Kumar Biswas, IIT Kharagpur

- **Quiz 11 Solution**
- **Practice Problems for week 12**

By

Arka Roy

NPTEL PMRF TA

Prime Minister's Research Fellow
Department of Electrical Engineering, IIT Patna
Web: https://sites.google.com/view/arka-roy/home

Powered by:

FREE ONLINE EDUCATION swayam
शिक्षित भारत, उन्नत भारत

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
भारतीय प्रौद्योगिकी संस्थान खड़गपुर
1951
योगः कर्मसु कौशलम्

PMRF
Prime Minister's Research Fellows
Ministry of Education
Government of India
सत्यमेव जयते

भारतीय प्रौद्योगिकी संस्थान पटना
INDIAN INSTITUTE OF TECHNOLOGY PATNA
विद्यार्थी लभते विद्यां

Which of following can be a target output of semantic segmentation problem with 4 class?

a.

0	1	0
0	1	0
1	0	0

I

1	1	0
0	0	0
0	0	0

II

0	0	1
1	0	0
0	0	0

III

0	0	0
1	0	0
0	0	0

IV

b.

0	1	0
0	1	0
1	0	0

I

1	0	0
0	1	0
0	0	0

II

0	0	1
1	0	0
0	0	0

III

0	1	0
0	0	0
0	1	1

IV

c.

0	1	0
0	1	0
1	0	0

I

1	0	0
0	0	1
0	0	0

II

0	0	1
1	0	0
0	0	0

III

0	0	0
0	0	0
0	1	1

IV

d.

0	1	0
0	1	0
1	1	0

I

1	0	0
1	0	0
0	1	0

II

0	0	1
1	0	0
0	0	1

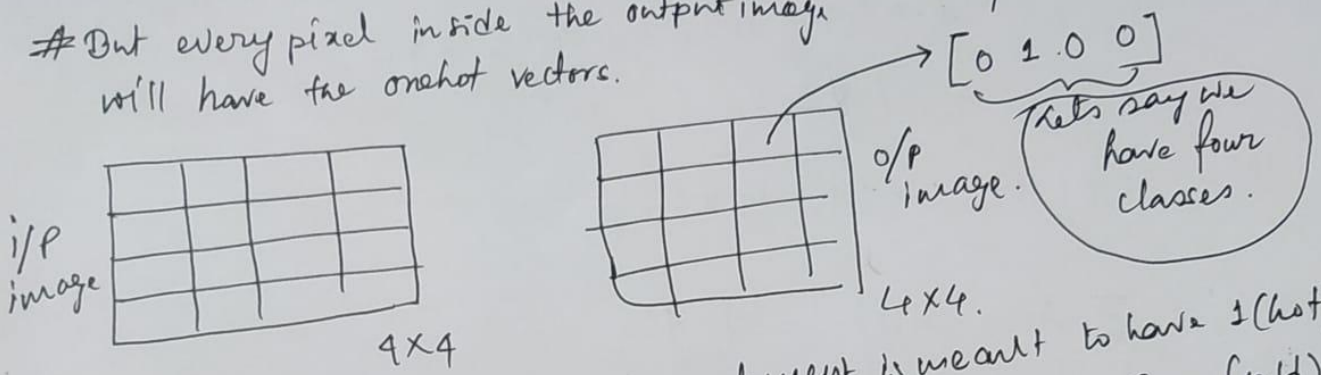
III

0	0	0
0	0	0
0	1	1

IV

① In case of image segmentation, the input image shape should be equal to output image shape.

But every pixel inside the output image will have the onehot vectors.



in case of onehot vector only one element is meant to have 1 (hot) rest are zero (cold)

Therefore in the options we have to search for that set every element in a particular pixel point should have the same property of one hot vector.

What will be the dice coefficient of following two one hot encoded vector? (|A|=no of 1 bit)

A	1	0	1	0	0	0	1	1	1	0	0	1	0	1
B	1	0	0	0	0	1	1	1	0	0	0	1	0	0

a. 0.83

b. 0.41

c. 0.67

d. 0.90

(2)

1	0	1	0	0	1	1	1	0	0	1	0	1	→ A
1	0	0	0	1	1	1	0	0	0	1	0	0	→ B.

|A| = no of 1 bit.

|A| = 7. |B| = 5.

|A∩B| = |Multiplication between A & B|

= |{1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0}|

= 4.

$$\text{Dice Coeff} = \frac{2|A \cap B|}{|A| + |B|}$$

$$= \frac{2 \times 4}{7+5} = \frac{2 \times 4}{12} = \frac{2}{3} = 0.67.$$

What will be the value of dice coefficient between A and B?

0.01	0.03	0.02	0.02
0.05	0.12	0.09	0.07
0.89	0.85	0.88	0.91
0.99	0.97	0.95	0.97

A

0	0	0	0
0	0	0	0
1	1	1	1
1	1	1	1

(Consider, |A|= sum of all elements

a. 0.23

b. 0.77

c. 0.11

d. 0.93

(3) $|A| = 0.01 + 0.03 + 0.02 + 0.02 + 0.05 + 0.12 + 0.09 + 0.07 + 0.89 + 0.85 + 0.88 + 0.91 + 0.99 + 0.97 + 0.95 + 0.97.$

$= 7.42$

$|B| = 1+1+1+1+1+1+1+1 = 8.$

$|A \cap B| = \sum_i A(i) \cdot B(i) = 7.42.$

Dice coeff $= \dfrac{2 \times |A \cap B|}{|A| + |B|}$

Suppose you have a $1D$ signal $x = [1,2,3,4,5]$ and a filter $f = [1,2,3,4]$, and you perform stride 2 transpose convolution on the signal x by the filter f to get the signal y. What will be the signal y if we don't perform cropping?

a. $y = [1,2,5,8,9,14,13,20,19,26,3,4]$
b. $y = [1,2,3,4,5,4,3,2,1]$
c. $y = [1,2,5,8,9,14,13,20,17,26,15,20]$

(4)

$x = [1 \quad 2 \quad 3 \quad 4 \quad 5]$

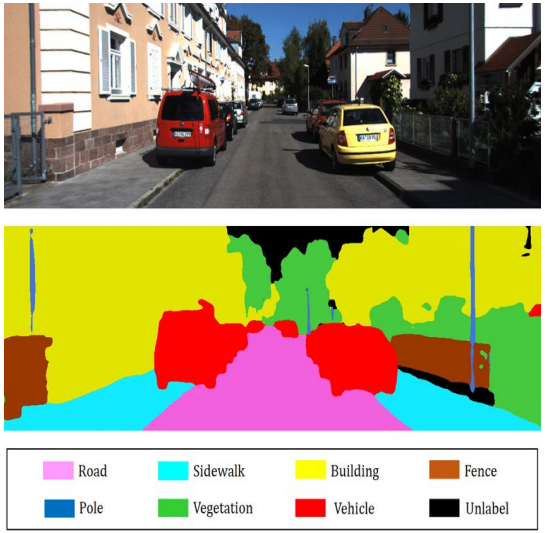
kernel

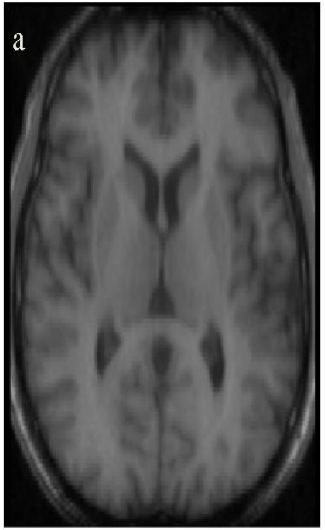
$f(0) = 1$
$f(1) = 1 \times 2 = 2$
$f(2) = 3 + 2 = 5$
$f(3) = 4 + 4 = 8$
$f(4) = 6 + 3 = 9$
$f(5) = 8 + 6 = 14$
$f(6) = 9 + 4 = 13.$
$f(7) = 12 + 8 = 20$
$f(8) = 12 + 5 = 17$
$f(9) = 16 + 10 = 26.$
$f(10) = 15$
$f(11) = 20$

$y = [1, 2, 5, 8, 9, 14, 13, 20, 17, 26, 15, 20].$

Which of the following is true for semantic segmentation?

a. Semantic Segmentation can be considered as pixel wise classification problem.
b. Semantic Segmented output has same dimension as the input image dimension.
c. It has application in Autonomous driving, Industrial inspection, and Medical imaging analysis.
d. All of the above





In a Deep CNN architecture, the feature map before applying a max pool layer with (2x2) kernel, stride 2 is given bellow.

2	30	3	14
14	12	7	10
4	1	14	19
2	5	16	2

After few successive convolution layers, the feature map is again up-sampled using Max Un-pooling. what will be the output of the Max-Unpooling layer?

a.

2	30	3	14
14	12	7	10
4	1	14	19
2	5	16	2

b.

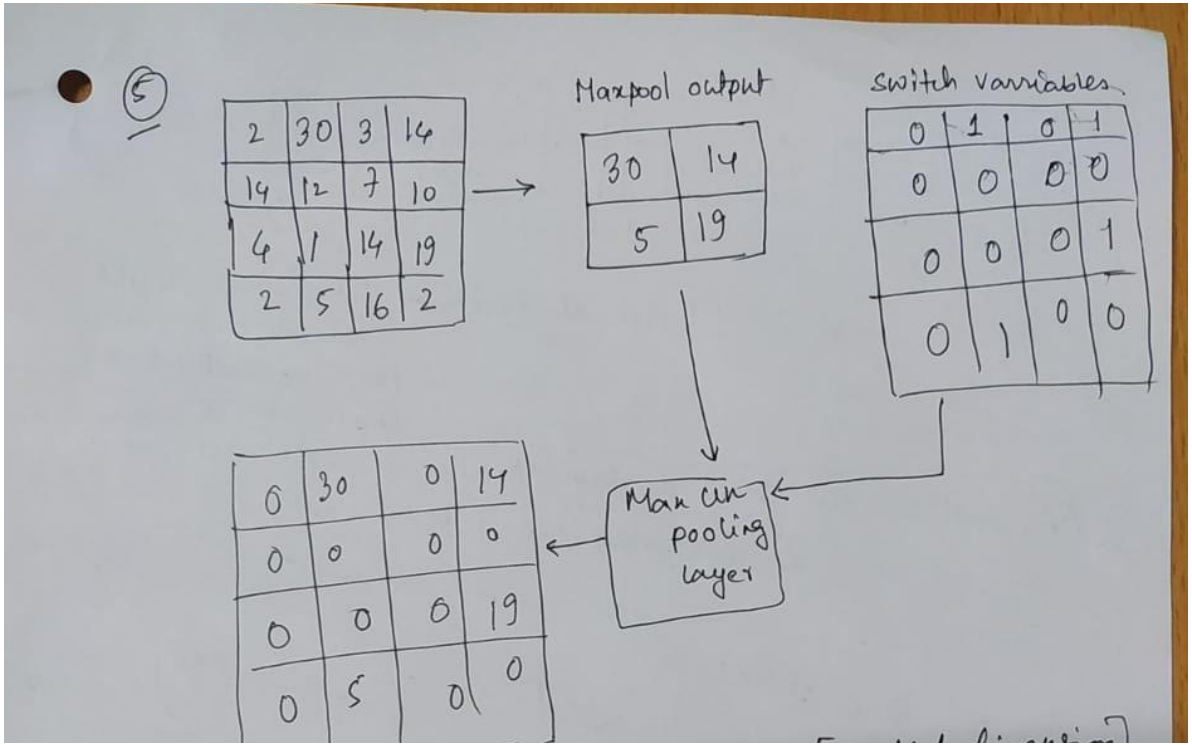
0	0	0	14
14	0	0	0
0	0	0	0
0	5	0	2

c.

0	30	0	14
0	0	0	0
0	0	0	19
0	5	0	0

d.

2	0	3	0
0	0	0	0
0	1	0	0
0	0	0	2



Which of the following operation reduces spatial dimension of features?

a. Max un-Pooling
b. Convolution with 3 × 3 Kernel, Stride=2, Padding all sides = 1
c. Convolution with 3 × 3 Kernel, Stride=1, Padding all sides = 1
d. Transposed convolution

(b) Lets take the image shape is: $x \times x$ [spatial dimension].

or $= \underbrace{x \times x}_{} \times 3$ → channel dimension.

(c) 3x3 Kernel, stride = 1, Pad = 1 :→

$$\text{o/p dimension} = \frac{x - \text{Kernel} + (2 \times \text{Padding})}{(\text{stride})} + 1$$

$$= \frac{x - 3 + 2}{1} + 1$$

i/pdim = o/Pdim

$$= x - 3 + 3 = x.$$

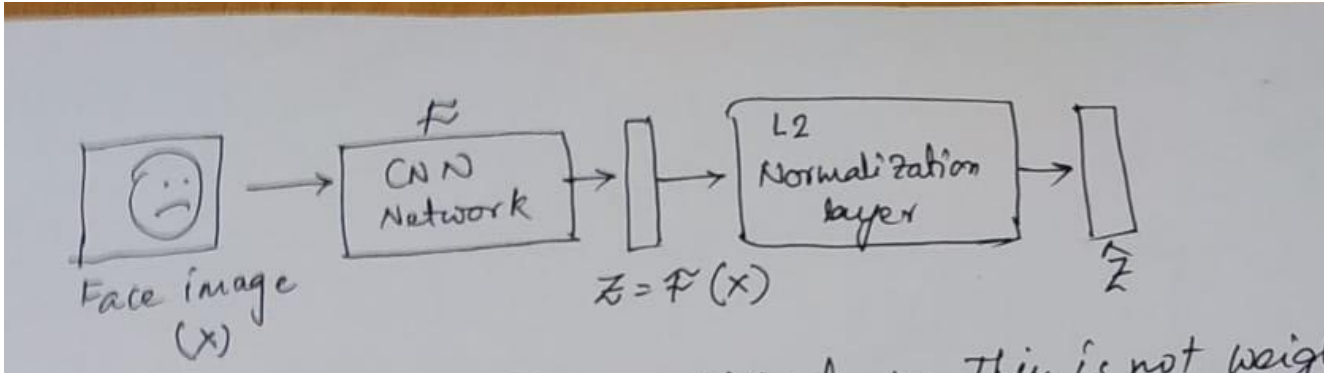
(b) stride = 2,

$$\text{o/Pdim} = \frac{x - 3 + 2 \times 1}{2} + 1$$

$$= \frac{x - 1}{2} + 1$$??

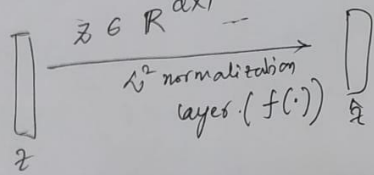
In FaceNet, why the L2 normalization layer is used?

a. To constrain the embedding function in a d-dimensional hyper-sphere.
b. For regularization of weight vector, i.e. L2 regularization.
c. For getting a sparse embedding function.
d. None of the above.



z is L^2 Normalized → This is not weight normalization as the normalization is applied on z

★ As well as this is not a Regularization of weight vector as it is not L^2 regularizer that is beingasked.



$$\hat{z} = f(z \in \mathbb{R}^{d \times 1}) = \sqrt{z_1^2 + z_2^2 + z_3^2 + \cdots + z_d^2} = 1.$$

$$\therefore \hat{z} = \sqrt{\sum_{\forall i} z_i^2} = 1.$$

$$\sum_{\forall i} z_i^2 = 1.$$

if it is a three dim vector then :→

$$z_1^2 + z_2^2 + z_3^2 = 1 \longrightarrow \text{sphere.}$$

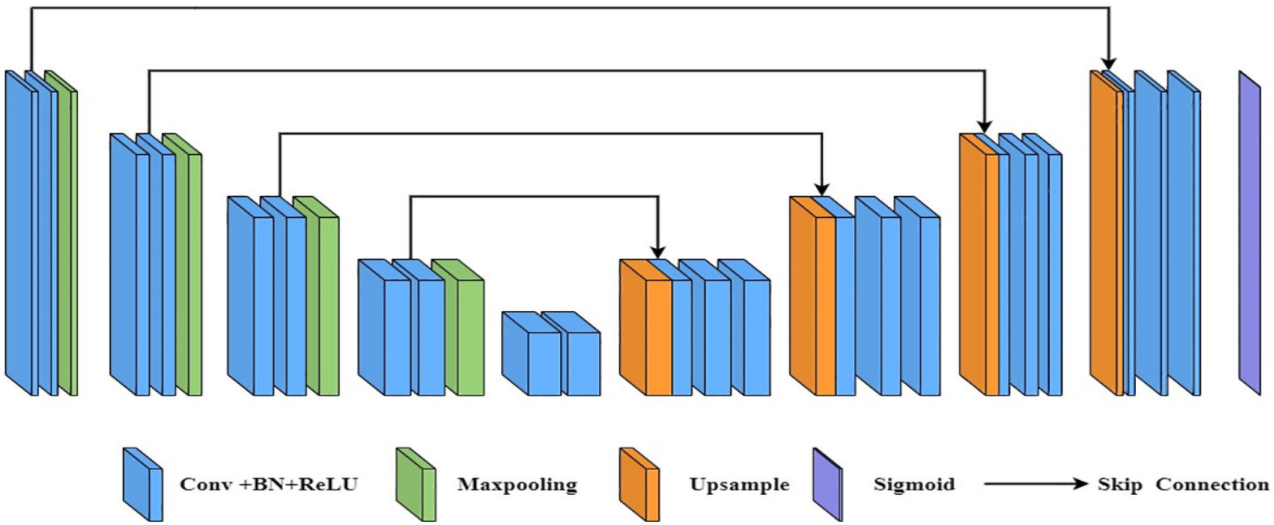
if $d > 3$ then $\underbrace{z_1^2 + z_2^2 + z_3^2 + \cdots = 1.}$

↳ High dimensional sphere.

↳ Hyper sphere.

What is the use of Skip Connection in image denoising networks?

a. Helping de-convolution layer to recover an improved clean version of image.
b. Back propagating the gradient to bottom layers, which makes the training easy.
c. To create the direct path between convolution layer and the corresponding mirror de-convolution layer.
d. All of the above.



What are the different challenges one face while creating a facial recognition system?

a. Different illumination condition
b. Different pose and orientation of face images
c. Limited dataset for training
d. All of the above

Learning Face Recognition from Limited Training Data using Deep Neural Networks

Xi Peng[1]
Department of Computer Science
Rutgers University
Piscataway, New Jersey 08854

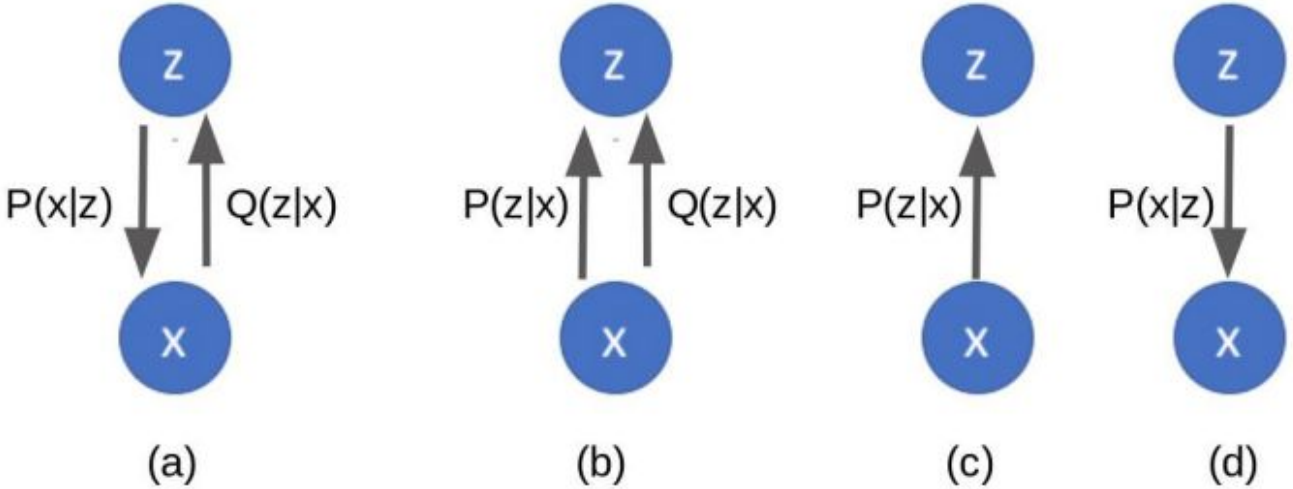
Nalini Ratha
IBM Thomas J. Watson
Research Center
Yorktown Heights, New York 10598

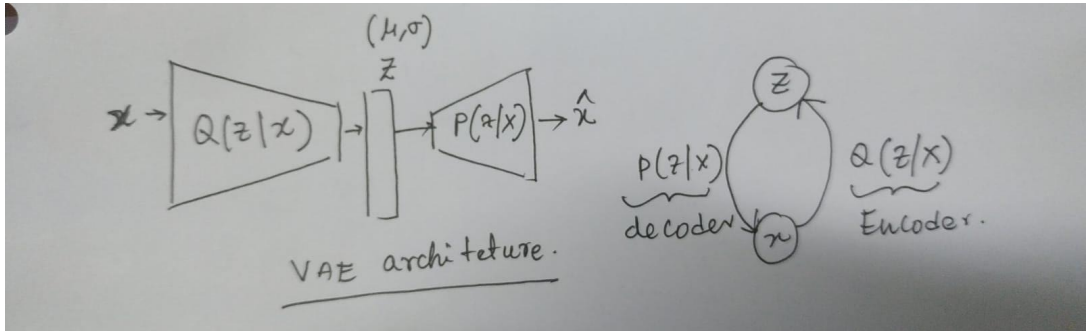
Sharathchandra Pankanti
IBM Thomas J. Watson
Research Center
Yorktown Heights, New York 10598

Multi-Angled Face Segmentation and Identification using Limited Data

Dane Brown
Department of Computer Science
Rhodes University
Grahamstown, South Africa
d.brown@ru.ac.za

Which one of the following graphical models fully represents a Variational Auto-encoder (VAE) realization?

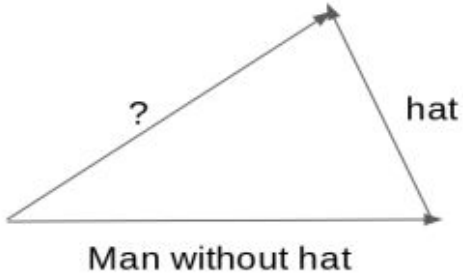




Which of the following is an INVALID activation function inside a neural network?

a. $f(x) = max(0, 2x)$
b. $f(x) = min(0, 2x)$
c. $f(x) = tanh(x)$
d. None of the above

Figure shows latent vector addition of two concepts of “man without a hat” and “hat”. What is expected from the resultant vector?



a. Hat without man
b. Man with hat
c. Woman with hat
d. Woman without hat