Course outline

About NPTEL

How does an NPTEL online course work?

Week 0

Week 1

Week 2

Week 3

Week 4

Week 5

Week 6

- CNNs for Object Detection: Pre-Deep Learning Era and Initial Steps
- CNNs for Object Detection: Two-Stage Models
- CNNs for Object Detection: Single-stage Models
- CNNs for Segmentation
- Lecture Materials
- Week 6 Feedback Form : Deep Learning for Computer Vision
- Practice: Week 6 : Assignment 6(Non-Graded)
- Quiz: Week 6: Assignment 6

Week 7

Download Videos

Text Transcripts

Books

Problem Solving Session - July 2024

Week 6: Assignment 6

The due date for submitting this assignment has passed.

Due on 2024-09-04, 23:59 IST.

Assignment submitted on 2024-09-04, 23:46 IST

1) Match the following: *1 point*

1) YOLO v1	i) Uses RepVGG-based backbone
2) YOLO v2	ii) CSPDarknet backbone
3) YOLO v3	iii) Use of Darknet-19
4) YOLO v4	iv) Feature pyramid networks
	v) Single grid cell prediction

○ 1⟶ v, 2⟶i, 3⟶iv, 4⟶ii

○ 1⟶iv, 2⟶i, 3⟶iii, 4⟶ii

○ 1⟶v, 2⟶iii, 3⟶i, 4⟶ii

● 1⟶v, 2⟶iii, 3⟶iv, 4⟶ii

Yes, the answer is correct.
Score: 1

Accepted Answers:
1⟶v, 2⟶iii, 3⟶iv, 4⟶ii

2) Match the following: *1 point*

1) VGGNet	i) 1×1 convolution
2) EfficientNet	ii) identity mapping
3) GoogleNet	iii) 3×3 convolution
4) ResNet	iv) 7×7 convolution
	v) depth-wise separable convolutions

○ 1⟶ iv, 2⟶ii, 3⟶v, 4⟶i

○ 1⟶iv, 2⟶iii, 3⟶v, 4⟶ii

○ 1⟶iv, 2⟶v, 3⟶ii, 4⟶iii

● 1⟶iii, 2⟶v, 3⟶i, 4⟶ii

Yes, the answer is correct.
Score: 1

Accepted Answers:
1⟶iii, 2⟶v, 3⟶i, 4⟶ii

3) The ConvNEXT model is an evolution of convolutional neural networks (CNNs), designed to bridge the performance gap with vision transformers (ViTs). One of the key architectural innovations in ConvNEXT is the introduction of depthwise convolution blocks inspired by transformers. Which of the following statements regarding the architectural design of ConvNEXT is incorrect? *1 point*

○ ConvNEXT employs depthwise convolutions followed by pointwise convolutions, similar to the MobileNet architecture, but with additional layer normalization and GELU activation.

● ConvNEXT replaces the traditional ResNet bottleneck block with a modified block that removes the ReLU activation function in favor of more non-linear operations like Swish.

○ The ConvNEXT model increases the size of the convolutional kernels to 7x7 to better capture long-range dependencies, mimicking the self-attention mechanism in transformers.

○ ConvNEXT introduces LayerNorm after the depthwise convolution and before the pointwise convolution to stabilize the training dynamics.

Yes, the answer is correct.
Score: 1

Accepted Answers:
ConvNEXT replaces the traditional ResNet bottleneck block with a modified block that removes the ReLU activation function in favor of more non-linear operations like Swish.

4) Consider an object detection system evaluated on a dataset consisting of 1000 images. The system makes 1500 predictions across these images, and for each image, there are annotated ground truth bounding boxes. The system's precision-recall curve is calculated, and the precision at different recall levels for one of the classes is as follows: *1 point*

Recall	Precision
0.1	0.90
0.2	0.85
0.3	0.80
0.4	0.75
0.5	0.70
0.6	0.65
0.7	0.60
0.8	0.55
0.9	0.50
1.0	0.45

Calculate the Average Precision (AP) for this class using the 11-point interpolation method, which averages the precision values at recall levels {0.0, 0.1, 0.2, ..., 1.0}. The precision at recall 0.0 can be assumed to be 1.0.
Additionally, the system's AP values for the other two classes are as follows:
- AP for class 2: 0.78
- AP for class 3: 0.72
Based on these AP values, what is the mean Average Precision (mAP) across all three classes?

○ 0.70

○ 0.69

● 0.73

○ 0.76

Yes, the answer is correct.
Score: 1

Accepted Answers:
0.73

5) Match the following computer vision tasks to situations: *1 point*

1) Instance Segmentation	i) There are dogs in these pixels
2) Classification	ii) There are 4 dogs in the image, and here are the pixels with the shape of each of their occurrence
3) Semantic Segmentation	iii) There is dog in image
4) Object Detection	iv) There are 4 dogs in the image

○ 1 ⟶ iv, 2 ⟶ iii, 3 ⟶ ii, 4 ⟶ i

● 1 ⟶ ii, 2 ⟶ iii, 3 ⟶ i, 4 ⟶ iv

○ 1 ⟶ iv, 2 ⟶ iii, 3 ⟶ i, 4 ⟶ ii

1 → iv, 2 → iii, 3 → i, 4 → ii

○ 1 → ii, 2 → iv, 3 → iii, 4 → i

Yes, the answer is correct.
Score: 1

Accepted Answers:
1 → ii, 2 → iii, 3 → i, 4 → iv

6) Which one of the following statements is false? ***1 point***

○ EfficientNet employs a compound scaling method to improve model efficiency and accuracy across different scales.
○ MobileNet utilizes depthwise separable convolutions to reduce the computational complexity of the network.
● DenseNet only connects each layer to the last layer in a feedforward fashion to address the vanishing gradient problem.
○ SeNet incorporates spatial squeeze-and-excitation blocks to enhance the representation power of the network by explicitly modeling channel-wise dependencies.

Yes, the answer is correct.
Score: 1

Accepted Answers:
DenseNet only connects each layer to the last layer in a feedforward fashion to address the vanishing gradient problem.

7) Which one of the following object detection networks uses an ROI pooling layer? ***1 point***

● Fast R-CNN
○ R-CNN
○ YOLO
○ All of the above

Yes, the answer is correct.
Score: 1

Accepted Answers:
Fast R-CNN

8) Consider two 12×12 bounding boxes (one on the upper left and one on the lower right) in an image with an overlapping region of 8 × 8. The Intersection over Union (IoU) score between the two boxes is (choose the closest value): ***1 point***

○ 10%
○ 18%
● 28%
○ 35%

Yes, the answer is correct.
Score: 1

Accepted Answers:
28%

9) The integral image of the image $\begin{bmatrix} 3 & 7 & 2 \\ 5 & 4 & 6 \\ 8 & 1 & 9 \end{bmatrix}$ ***1 point***

○ $\begin{bmatrix} 3 & 6 & 9 \\ 12 & 15 & 18 \\ 21 & 24 & 27 \end{bmatrix}$

● $\begin{bmatrix} 3 & 10 & 12 \\ 8 & 19 & 27 \\ 16 & 28 & 45 \end{bmatrix}$

○ $\begin{bmatrix} 3 & 7 & 12 \\ 8 & 15 & 25 \\ 16 & 28 & 45 \end{bmatrix}$

○ $\begin{bmatrix} 9 & 6 & 3 \\ 8 & 5 & 2 \\ 7 & 4 & 1 \end{bmatrix}$

Yes, the answer is correct.
Score: 1

Accepted Answers:
$\begin{bmatrix} 3 & 10 & 12 \\ 8 & 19 & 27 \\ 16 & 28 & 45 \end{bmatrix}$

10) Which of the following is true? Select all possible answers: ***1 point***

□ The size of the effective receptive field reduces as we go deeper in Convolution Neural Network.
☑ For transfer learning, if the source and target datasets are dissimilar and the target dataset size is small, it is better to only use source task model's general layers and append a new classifier to it instead of tuning on source task model's specific layers
□ The number of FLOPS required by EfficientNet*B6* is less than the number of FLOPS required by EfficientNet*B3*
□ When using 1 × 1 convolution on a feature map, it is good practice to apply padding since 1 × 1 reduces the height and width of feature map.
☑ In 3D convolution, the kernel moves in 3 directions and the input data is 4-dimensional

Yes, the answer is correct.
Score: 1

Accepted Answers:
For transfer learning, if the source and target datasets are dissimilar and the target dataset size is small, it is better to only use source task model's general layers and append a new classifier to it instead of tuning on source task model's specific layers
In 3D convolution, the kernel moves in 3 directions and the input data is 4-dimensional

Let input have size $D_f \times D_f \times M$ where $D_f = 128$ and $M = 16$ and output feature map (after passing input through conv layer) has $D_f \times D_f \times N$ size where $N = 32$. Assume padded convolution. Let width of the square kernel in conv layer be k where $k = 5$ (Ignore the bias term in the calculation).

Calculate the number of parameters and computational cost for this convolution layer.

11) Number of Parameters:

12800

Yes, the answer is correct.
Score: 0.5

Accepted Answers:
(Type: Numeric) 12800

0.5 points

12) Computational Cost:

209715200

Yes, the answer is correct.
Score: 0.5

Accepted Answers:
(Type: Numeric) 209715200

0.5 points

Using the same dimensions specified in the previous question, calculate the number of parameters and computational cost, but make use of **Depthwise Seperable convolution** instead of standard convolution.

13) Number of parameters for depthwise convolution:

400

Yes, the answer is correct.
Score: 0.25

Accepted Answers:
(Type: Numeric) 400

0.25 points

14) Computational Cost for for depthwise convolution:

6553600

Yes, the answer is correct.
Score: 0.25

Accepted Answers:
(Type: Numeric) 6553600

0.25 points

15) Number of parameters for pointwise convolution:

512

Yes, the answer is correct.
Score: 0.25

Accepted Answers:
(Type: Numeric) 512

0.25 points

16) Computational cost for for pointwise convolution:

8388608

Yes, the answer is correct.
Score: 0.25

Accepted Answers:
(Type: Numeric) 8388608

0.25 points