ARTIFICIAL INTELLIGENCE - UE15CS325

END SEMESTER ASSESSMENT

VARUN KUMAR S - 01FB15ECS 338

VARUN V - 01FB15ECS 341

VARUN Y VORA - 01FB15ECS 342

VISHWAS S - 01FB15ECS 355

IMAGE CLASSIFICATION AND LOCALISATION FOR PASCAL VOC 2010

1) DATASET Preparation The size of the dataset is 4474. It was randomly shuffled and was divided as follows: Training Size : 2500 Validation Size : 1000 Evaluation Size : 1244

Each image was preprocessed using 'pillow' to reduce its size and padded with 0's This made the dimension of each image (128, 128, 3) The bounding boxes were also modified accordingly

The 15 classes were converted to one-hot vectors. All listed metrics are for top prediction only.

2) Building CLASSIFIER

2 convolution layers along with pooling was used. 2 Dense layers were stacked on top of this. A dropout was added for all dense layers. This was followed by a softmax layer. We directly take the highest value as our prediction.

Accuracy : 35.93%

3) Object LOCALISATION

2 convolution layers along with pooling was used. 3 Dense Layers with relu activation are stacked on top. The final dense layer gives 4 predictions- xmin, ymin, xmax, ymax.

Accuracy : 63.86%

4) Classifier using Transfer Learning

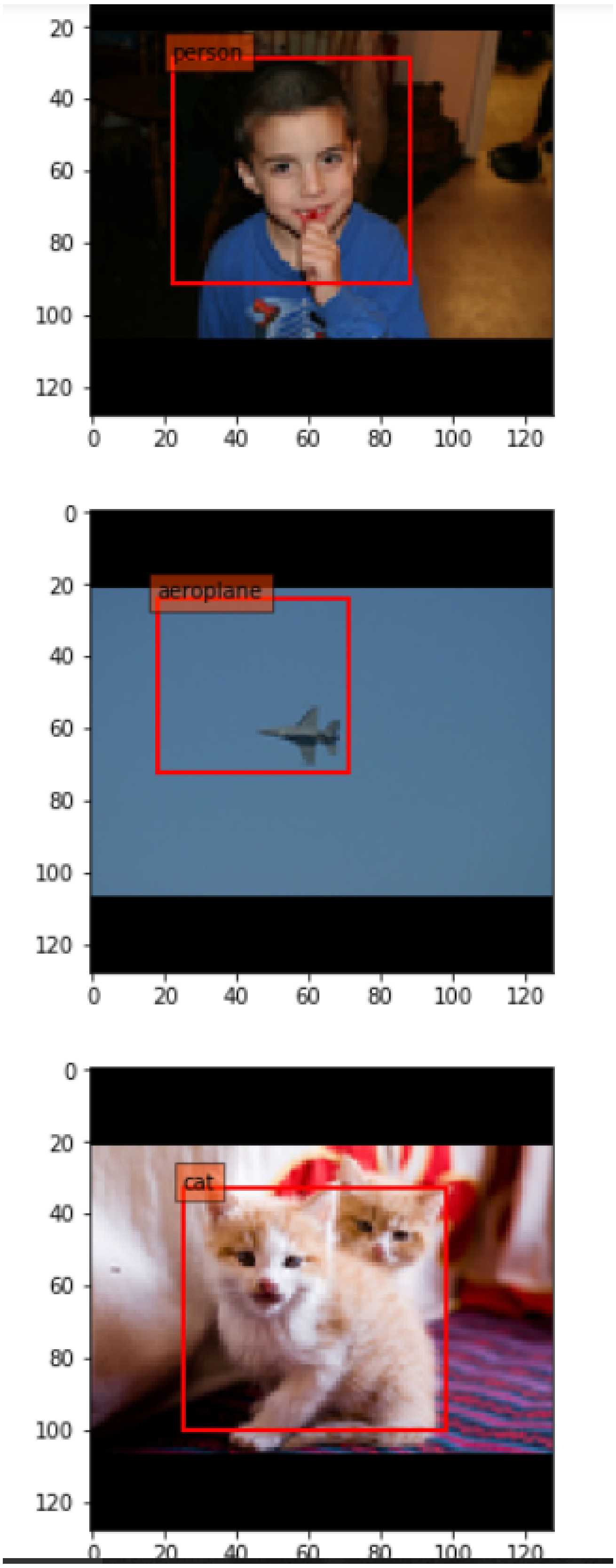
VGG16 model with weights trained in Imagenet was used. We removed the top 2 layers and added 2 Dense Layers with Relu and softmax activations.

Accuracy : 58.5 %

5) Localisation using Transfer Learning

VGG16 model was used without its top 2 layers. 3 Dense Layers with 0.5 Dropout. However, a large gain in performance was not observed. Accuracy : 65.54%

Object Detection



In [18]:

```
from keras.models import load_model

classifier = load_model('classifier_model.h5')
print('\nCLASSIFICATION\n')
classifier.summary()

localisation = load_model('box_model.h5')
print('\nLOCALISAT
```

In [18]:

```
ION\n')
localisation.summary()

classifer_transfer = load_model('transfer_learning_model.h5')
print('\nCLASSIFICATION USING TRANSFER LEARNING\n')
classifer_transfer.summary()

localisation_transfer = load_model('')
print('\nLOCALISATION USING TRANSFER LEARNING\n')
localisation_transfer.summary()
```

```
CLASSIFICATION

_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
conv2d_9 (Conv2D)            (None, 126, 126, 64)      1792      
_________________________________________________________________
max_pooling2d_9 (MaxPooling2 (None, 31, 31, 64)        0         
_________________________________________________________________
conv2d_10 (Conv2D)           (None, 29, 29, 128)       73856     
_________________________________________________________________
max_pooling2d_10 (MaxPooling (None, 7, 7, 128)         0         
_________________________________________________________________
flatten_5 (Flatten)          (None, 6272)              0         
_________________________________________________________________
dropout_12 (Dropout)         (None, 6272)              0         
_________________________________________________________________
dense_13 (Dense)             (None, 128)               802944    
_________________________________________________________________
dropout_13 (Dropout)         (None, 128)               0         
_________________________________________________________________
dense_14 (Dense)             (None, 64)                8256      
_________________________________________________________________
dense_15 (Dense)             (None, 15)                975       
=================================================================
Total params: 887,823
Trainable params: 887,823
Non-trainable params: 0
_________________________________________________________________

LOCALISATION

_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
conv2d_5 (Conv2D)            (None, 126, 126, 32)      896       
_________________________________________________________________
max_pooling2d_5 (MaxPooling2 (None, 31, 31, 32)        0         
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 29, 29, 64)        18496     
_________________________________________________________________
max_pooling2d_6 (MaxPooling2 (None, 7, 7, 64)          0         
_________________________________________________________________
flatten_3 (Flatten)          (None, 3136)              0         
_________________________________________________________________
dropout_7 (Dropout)          (None, 3136)              0         
_________________________________________________________________
dense_7 (Dense)              (None, 64)                200768    
```

```
dense_7 (Dense)              (None, 64)                200768
_________________________________________________________________
dropout_8 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_8 (Dense)              (None, 32)                2080
_________________________________________________________________
dropout_9 (Dropout)          (None, 32)                0
_________________________________________________________________
dense_9 (Dense)              (None, 4)                 132
=================================================================
Total params: 222,372
Trainable params: 222,372
Non-trainable params: 0
_________________________________________________________________

CLASSIFICATION USING TRANSFER LEARNING

_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_4 (InputLayer)         (None, None, None, 3)     0
_________________________________________________________________
block1_conv1 (Conv2D)        (None, None, None, 64)    1792
_________________________________________________________________
block1_conv2 (Conv2D)        (None, None, None, 64)    36928
_________________________________________________________________
block1_pool (MaxPooling2D)   (None, None, None, 64)    0
_________________________________________________________________
block2_conv1 (Conv2D)        (None, None, None, 128)   73856
_________________________________________________________________
block2_conv2 (Conv2D)        (None, None, None, 128)   147584
_________________________________________________________________
block2_pool (MaxPooling2D)   (None, None, None, 128)   0
_________________________________________________________________
block3_conv1 (Conv2D)        (None, None, None, 256)   295168
_________________________________________________________________
block3_conv2 (Conv2D)        (None, None, None, 256)   590080
_________________________________________________________________
block3_conv3 (Conv2D)        (None, None, None, 256)   590080
_________________________________________________________________
block3_pool (MaxPooling2D)   (None, None, None, 256)   0
_________________________________________________________________
block4_conv1 (Conv2D)        (None, None, None, 512)   1180160
_________________________________________________________________
block4_conv2 (Conv2D)        (None, None, None, 512)   2359808
_________________________________________________________________
block4_conv3 (Conv2D)        (None, None, None, 512)   2359808
_________________________________________________________________
block4_pool (MaxPooling2D)   (None, None, None, 512)   0
_________________________________________________________________
block5_conv1 (Conv2D)        (None, None, None, 512)   2359808
_________________________________________________________________
block5_conv2 (Conv2D)        (None, None, None, 512)   2359808
_________________________________________________________________
block5_conv3 (Conv2D)        (None, None, None, 512)   2359808
_________________________________________________________________
block5_pool (MaxPooling2D)   (None, None, None, 512)   0
_________________________________________________________________
global_average_pooling2d_4 ( (None, 512)               0
_________________________________________________________________
dense_7 (Dense)              (None, 128)               65664
```

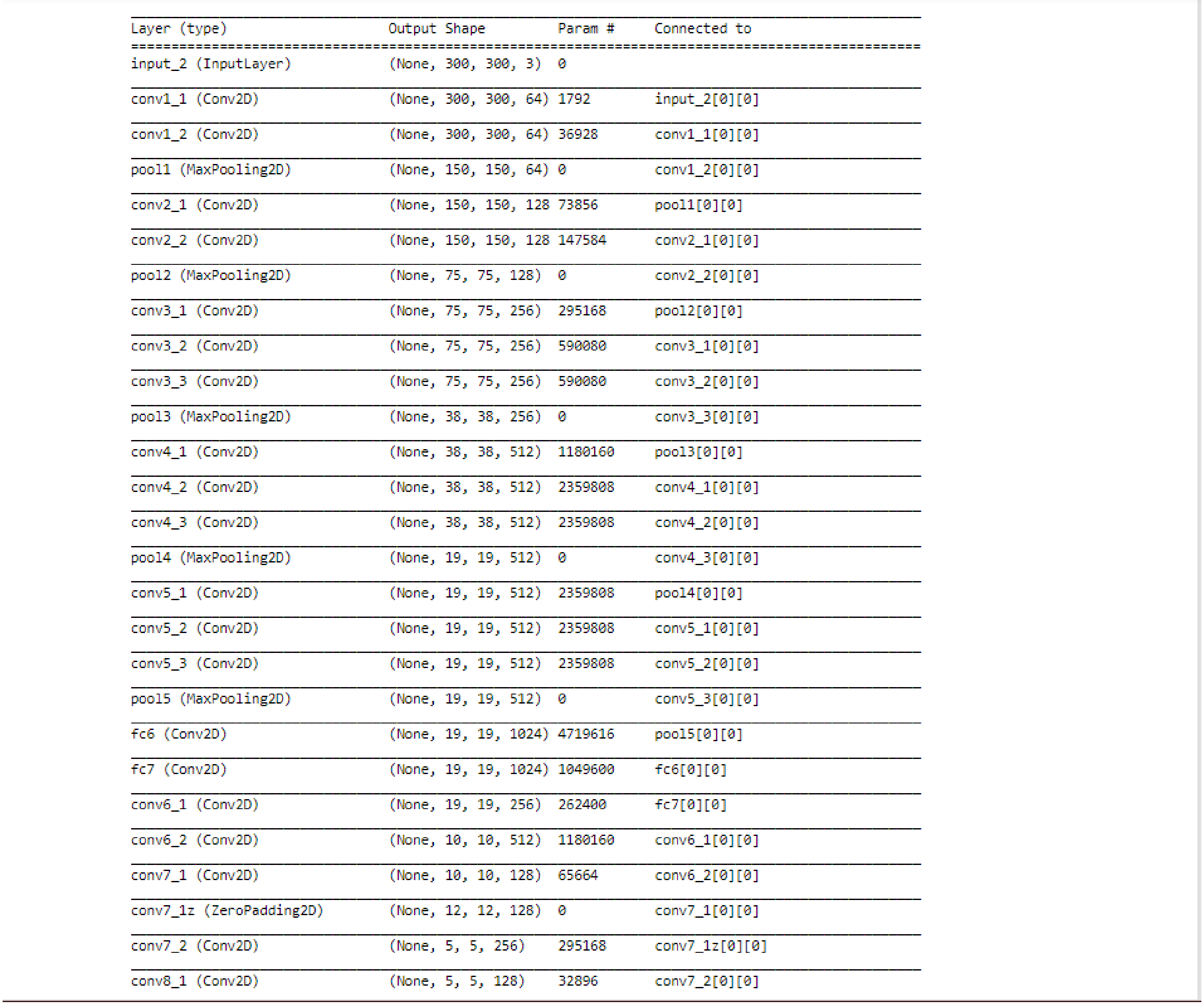
```
_________________________________________________________________
dense_8 (Dense)              (None, 15)                1935
=================================================================
Total params: 14,782,287
Trainable params: 67,599
Non-trainable params: 14,714,688
_________________________________________________________________
```

In [6]:

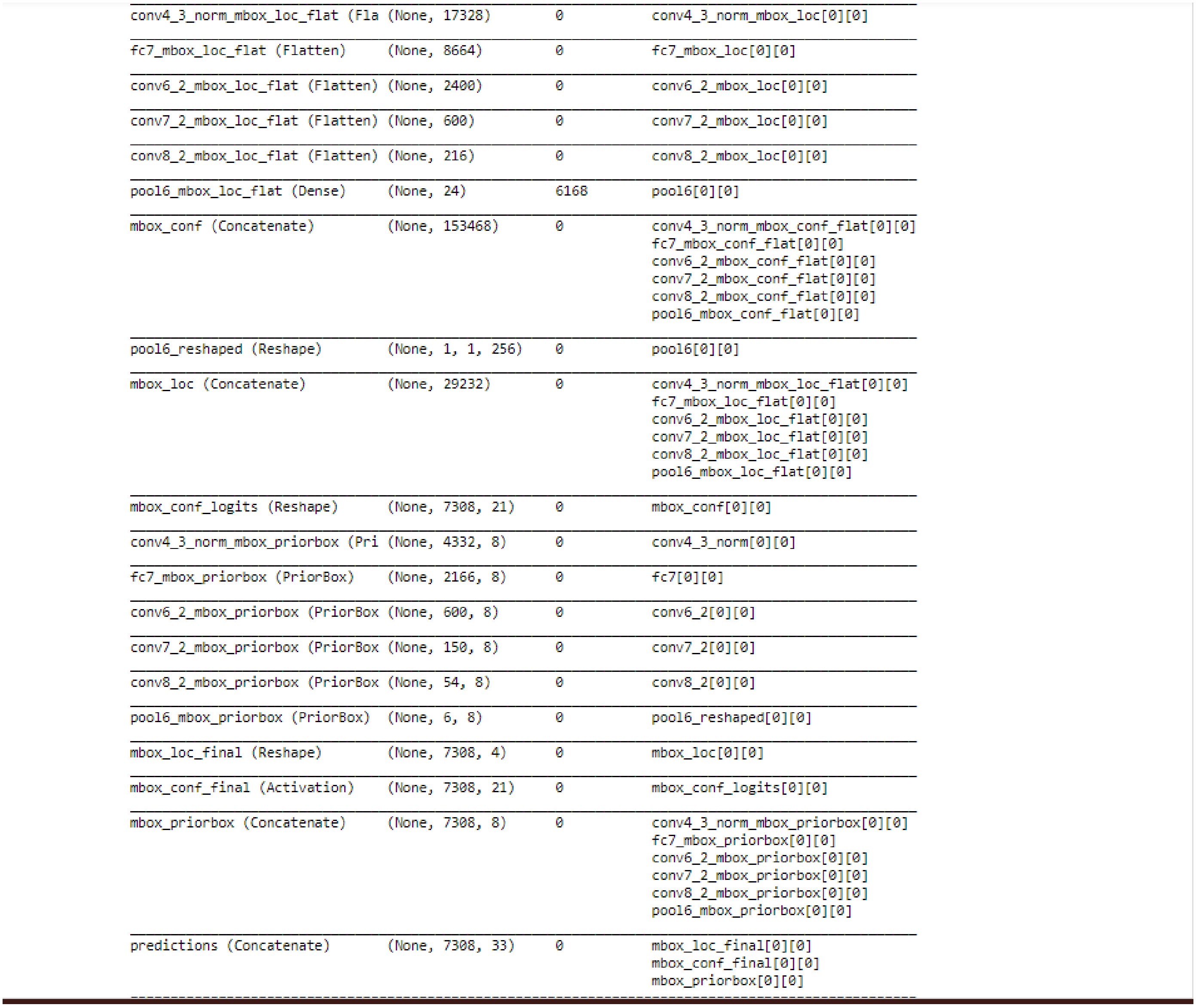
```
from NNfunctions import get_dataset, make_confusion_matrix
x, y, z = get_dataset('C:/Users/Varun/Desktop/AI-ESA/preprocessing')
```

```
[4744, 4744, 4744]
```

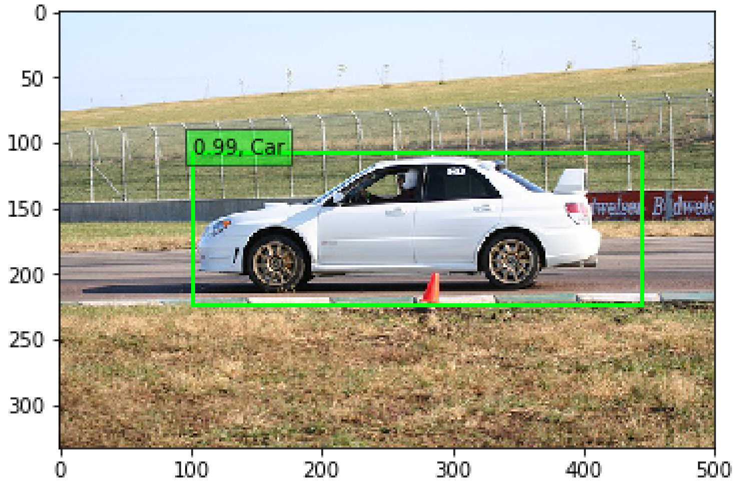
Object Detection and classification using TRANSFER LEARNING

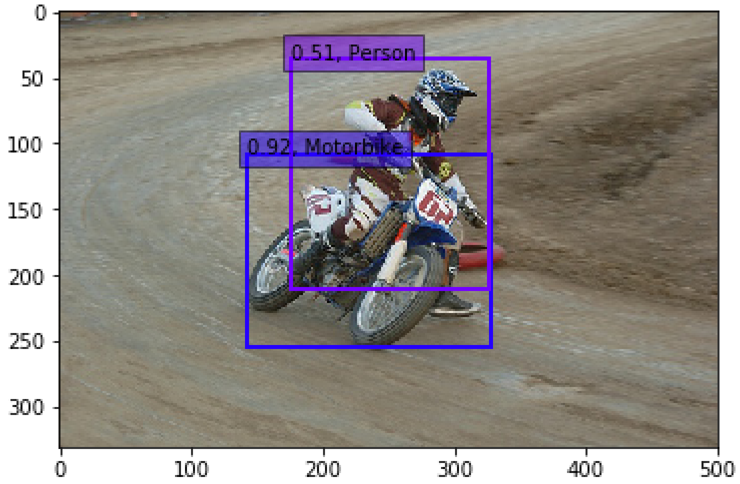


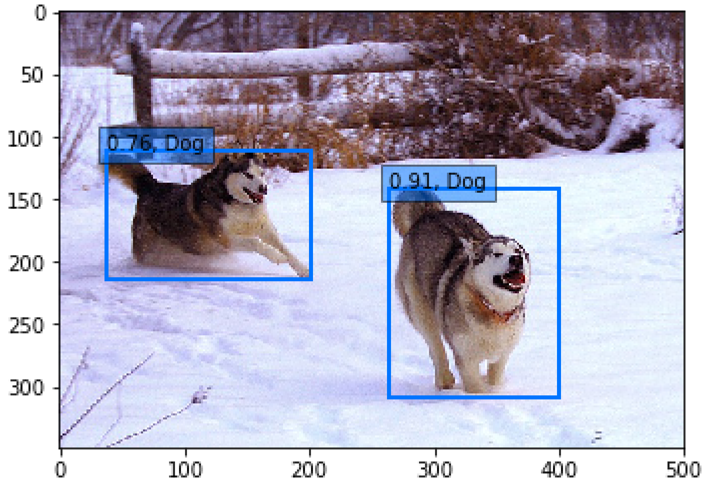
```
____________________________________________________________________________________________________
Layer (type)                     Output Shape          Param #     Connected to
====================================================================================================
input_2 (InputLayer)             (None, 300, 300, 3)   0
____________________________________________________________________________________________________
conv1_1 (Conv2D)                 (None, 300, 300, 64)  1792        input_2[0][0]
____________________________________________________________________________________________________
conv1_2 (Conv2D)                 (None, 300, 300, 64)  36928       conv1_1[0][0]
____________________________________________________________________________________________________
pool1 (MaxPooling2D)             (None, 150, 150, 64)  0           conv1_2[0][0]
____________________________________________________________________________________________________
conv2_1 (Conv2D)                 (None, 150, 150, 128  73856       pool1[0][0]
____________________________________________________________________________________________________
conv2_2 (Conv2D)                 (None, 150, 150, 128  147584      conv2_1[0][0]
____________________________________________________________________________________________________
pool2 (MaxPooling2D)             (None, 75, 75, 128)   0           conv2_2[0][0]
____________________________________________________________________________________________________
conv3_1 (Conv2D)                 (None, 75, 75, 256)   295168      pool2[0][0]
____________________________________________________________________________________________________
conv3_2 (Conv2D)                 (None, 75, 75, 256)   590080      conv3_1[0][0]
____________________________________________________________________________________________________
conv3_3 (Conv2D)                 (None, 75, 75, 256)   590080      conv3_2[0][0]
____________________________________________________________________________________________________
pool3 (MaxPooling2D)             (None, 38, 38, 256)   0           conv3_3[0][0]
____________________________________________________________________________________________________
conv4_1 (Conv2D)                 (None, 38, 38, 512)   1180160     pool3[0][0]
____________________________________________________________________________________________________
conv4_2 (Conv2D)                 (None, 38, 38, 512)   2359808     conv4_1[0][0]
____________________________________________________________________________________________________
conv4_3 (Conv2D)                 (None, 38, 38, 512)   2359808     conv4_2[0][0]
____________________________________________________________________________________________________
pool4 (MaxPooling2D)             (None, 19, 19, 512)   0           conv4_3[0][0]
____________________________________________________________________________________________________
conv5_1 (Conv2D)                 (None, 19, 19, 512)   2359808     pool4[0][0]
____________________________________________________________________________________________________
conv5_2 (Conv2D)                 (None, 19, 19, 512)   2359808     conv5_1[0][0]
____________________________________________________________________________________________________
conv5_3 (Conv2D)                 (None, 19, 19, 512)   2359808     conv5_2[0][0]
____________________________________________________________________________________________________
pool5 (MaxPooling2D)             (None, 19, 19, 512)   0           conv5_3[0][0]
____________________________________________________________________________________________________
fc6 (Conv2D)                     (None, 19, 19, 1024)  4719616     pool5[0][0]
____________________________________________________________________________________________________
fc7 (Conv2D)                     (None, 19, 19, 1024)  1049600     fc6[0][0]
____________________________________________________________________________________________________
conv6_1 (Conv2D)                 (None, 19, 19, 256)   262400      fc7[0][0]
____________________________________________________________________________________________________
conv6_2 (Conv2D)                 (None, 10, 10, 512)   1180160     conv6_1[0][0]
____________________________________________________________________________________________________
conv7_1 (Conv2D)                 (None, 10, 10, 128)   65664       conv6_2[0][0]
____________________________________________________________________________________________________
conv7_1z (ZeroPadding2D)         (None, 12, 12, 128)   0           conv7_1[0][0]
____________________________________________________________________________________________________
conv7_2 (Conv2D)                 (None, 5, 5, 256)     295168      conv7_1z[0][0]
____________________________________________________________________________________________________
conv8_1 (Conv2D)                 (None, 5, 5, 128)     32896       conv7_2[0][0]
```



```
__________________________________________________________________________________________________
conv4_3_norm_mbox_loc_flat (Fla (None, 17328)        0           conv4_3_norm_mbox_loc[0][0]
__________________________________________________________________________________________________
fc7_mbox_loc_flat (Flatten)     (None, 8664)         0           fc7_mbox_loc[0][0]
__________________________________________________________________________________________________
conv6_2_mbox_loc_flat (Flatten) (None, 2400)         0           conv6_2_mbox_loc[0][0]
__________________________________________________________________________________________________
conv7_2_mbox_loc_flat (Flatten) (None, 600)          0           conv7_2_mbox_loc[0][0]
__________________________________________________________________________________________________
conv8_2_mbox_loc_flat (Flatten) (None, 216)          0           conv8_2_mbox_loc[0][0]
__________________________________________________________________________________________________
pool6_mbox_loc_flat (Dense)     (None, 24)           6168        pool6[0][0]
__________________________________________________________________________________________________
mbox_conf (Concatenate)         (None, 153468)       0           conv4_3_norm_mbox_conf_flat[0][0]
                                                                 fc7_mbox_conf_flat[0][0]
                                                                 conv6_2_mbox_conf_flat[0][0]
                                                                 conv7_2_mbox_conf_flat[0][0]
                                                                 conv8_2_mbox_conf_flat[0][0]
                                                                 pool6_mbox_conf_flat[0][0]
__________________________________________________________________________________________________
pool6_reshaped (Reshape)        (None, 1, 1, 256)    0           pool6[0][0]
__________________________________________________________________________________________________
mbox_loc (Concatenate)          (None, 29232)        0           conv4_3_norm_mbox_loc_flat[0][0]
                                                                 fc7_mbox_loc_flat[0][0]
                                                                 conv6_2_mbox_loc_flat[0][0]
                                                                 conv7_2_mbox_loc_flat[0][0]
                                                                 conv8_2_mbox_loc_flat[0][0]
                                                                 pool6_mbox_loc_flat[0][0]
__________________________________________________________________________________________________
mbox_conf_logits (Reshape)      (None, 7308, 21)     0           mbox_conf[0][0]
__________________________________________________________________________________________________
conv4_3_norm_mbox_priorbox (Pri (None, 4332, 8)      0           conv4_3_norm[0][0]
__________________________________________________________________________________________________
fc7_mbox_priorbox (PriorBox)    (None, 2166, 8)      0           fc7[0][0]
__________________________________________________________________________________________________
conv6_2_mbox_priorbox (PriorBox (None, 600, 8)       0           conv6_2[0][0]
__________________________________________________________________________________________________
conv7_2_mbox_priorbox (PriorBox (None, 150, 8)       0           conv7_2[0][0]
__________________________________________________________________________________________________
conv8_2_mbox_priorbox (PriorBox (None, 54, 8)        0           conv8_2[0][0]
__________________________________________________________________________________________________
pool6_mbox_priorbox (PriorBox)  (None, 6, 8)         0           pool6_reshaped[0][0]
__________________________________________________________________________________________________
mbox_loc_final (Reshape)        (None, 7308, 4)      0           mbox_loc[0][0]
__________________________________________________________________________________________________
mbox_conf_final (Activation)    (None, 7308, 21)     0           mbox_conf_logits[0][0]
__________________________________________________________________________________________________
mbox_priorbox (Concatenate)     (None, 7308, 8)      0           conv4_3_norm_mbox_priorbox[0][0]
                                                                 fc7_mbox_priorbox[0][0]
                                                                 conv6_2_mbox_priorbox[0][0]
                                                                 conv7_2_mbox_priorbox[0][0]
                                                                 conv8_2_mbox_priorbox[0][0]
                                                                 pool6_mbox_priorbox[0][0]
__________________________________________________________________________________________________
predictions (Concatenate)       (None, 7308, 33)     0           mbox_loc_final[0][0]
                                                                 mbox_conf_final[0][0]
                                                                 mbox_priorbox[0][0]
==================================================================================================
```


0.99, Car
0
50
100
150
200
250
300
0
100
200
300
400
500


0.51, Person
0.92, Motorbike
0
50
100
150
200
250
300
0
100
200
300
400
500


0.76, Dog
0.91, Dog
0
50
100
150
200
250
300
0
100
200
300
400
500

In [17]:

```
from random import randint
k = randint(0,3744)
print('Confusion Matrix for Classification ')
make_confusion_matrix(classifier, x[k:k+500], z[k:k+500])
```

```
Confusion Matrix for Classification
CONFUSION MATRIX
[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14]
[[59  0  2  1  1  0  0 10 12  0  0  1  0  0  0]
 [ 4  3  0  1  0  0  0  1  0  0  7  0  0  0  0]
 [19  0 16  0  7  0  0  2  8  0  1  0  1  0  0]
 [10  0  0 11  6  0  0  1  0  1  2  0  0  0  0]
 [ 0  0  2  5 22  0  0  0  1  1  0  0  0  0  0]
 [ 4  0  0  3  0  0  0  1  2  0  0  0  0  0  0]
 [ 5  0  1  0  0  0  0  2  3  0  0  0  0  0  0]
 [22  0  0  0  1  0  0 43 16  1  1  0  0  0  0]
 [22  0  2  1  2  0  0 14 32  0  1  1  0  0  0]
 [ 1  0  0  3 13  0  0  0  0  1  2  0  0  0  0]
 [11  0  2  7  4  0  0  2  0  0  9  0  0  0  0]
 [ 5  0  1  3  0  0  0  0  0  0  0  1  0  0  0]
 [ 8  0  0  2  0  0  0  0  0  0  0  1  3  0  0]
 [ 4  0  3  1  0  0  0  2  2  0  0  0  0  2  0]
 [ 8  0  0  0  0  0  0  1  0  0  0  0  0  0  0]]
{0: 'person', 1: 'tvmonitor', 2: 'bird', 3: 'train', 4: 'aeroplane', 5: 'hor
se', 6: 'cow', 7: 'cat', 8: 'dog', 9: 'boat', 10: 'car', 11: 'bicycle', 12:
'motorbike', 13: 'bottle', 14: 'chair'}
[['ACCURACY' 'PRECISION' 'RECALL']
 ['0.3242' '0.6860' '0.2921']
 ['1.0000' '0.1875' '0.0149']
 ['0.5517' '0.2963' '0.0792']
 ['0.2895' '0.3548' '0.0545']
 ['0.3929' '0.7097' '0.1089']
 ['0.0000' '0.0000' '0.0000']
 ['0.0000' '0.0000' '0.0000']
 ['0.5443' '0.5119' '0.2129']
 ['0.4211' '0.4267' '0.1584']
 ['0.2500' '0.0500' '0.0050']
 ['0.3913' '0.2571' '0.0446']
 ['0.2500' '0.1000' '0.0050']
 ['0.7500' '0.2143' '0.0149']
 ['1.0000' '0.1429' '0.0099']
 ['0.0000' '0.0000' '0.0000']]
```

In [14]:

```
from random import randint
k = randint(0,3744)
make_confusion_matrix(classifer_transfer, x[k:k+500], z[k:k+500])
#Confusion matrix for a subset of the dataset
```

```
CONFUSION MATRIX
[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14]
[[74  1  1  4  3  2  0  0  1  0  0  0  0  0  1]
 [ 6 14  0  1  0  0  0  0  0  0  0  0  0  0  0]
 [ 7  0 43  1  3  0  0  9  8  0  0  0  0  0  1]
 [ 0  1  0 22  0  0  0  0  0  0  0  0  1  0  0]
 [ 1  0  1  6 30  0  0  1  2  0  1  0  0  0  0]
 [ 4  0  0  2  1  6  0  5  1  0  0  0  0  0  0]
 [ 2  0  0  0  0  0  0  1  7  0  0  0  0  0  0]
 [ 7  2  3  0  0  0  0 47 17  0  0  0  0  0  1]
 [ 9  0  5  1  1  0  0  7 44  0  0  0  0  0  0]
 [ 0  0  0  9  4  0  0  0  1  0  0  0  0  0  0]
 [ 1  0  1  4  4  0  0  0  2  0  9  0  0  0  0]
 [ 3  0  0  3  1  0  0  1  0  0  0  5  1  0  0]
 [ 4  0  0  0  0  0  0  0  1  0  1  2  4  0  0]
 [ 4  0  0  1  0  0  0  2  0  0  0  0  0  4  0]
 [ 3  0  1  2  0  0  0  0  0  0  0  0  0  1  2]]
{0: 'person', 1: 'tvmonitor', 2: 'bird', 3: 'train', 4: 'aeroplane', 5: 'hor
se', 6: 'cow', 7: 'cat', 8: 'dog', 9: 'boat', 10: 'car', 11: 'bicycle', 12:
'motorbike', 13: 'bottle', 14: 'chair'}
[['ACCURACY' 'PRECISION' 'RECALL']
 ['0.5920' '0.8506' '0.2434']
 ['0.7778' '0.6667' '0.0461']
 ['0.7818' '0.5972' '0.1414']
 ['0.3929' '0.9167' '0.0724']
 ['0.6383' '0.7143' '0.0987']
 ['0.7500' '0.3158' '0.0197']
 ['0.0000' '0.0000' '0.0000']
 ['0.6438' '0.6104' '0.1546']
 ['0.5238' '0.6567' '0.1447']
 ['0.0000' '0.0000' '0.0000']
 ['0.8182' '0.4286' '0.0296']
 ['0.7143' '0.3571' '0.0164']
 ['0.6667' '0.3333' '0.0132']
 ['0.8000' '0.3636' '0.0132']
 ['0.4000' '0.2222' '0.0066']]
```

Observations and Conclusion:

1. The metrics for car show that the model is able to classify an object as a car relatively well.
2. The most confusion is between cat and dog.
3. The bounded boxes appear to around the same region for most predictions. Probably the model has learnt where the image most likely appears.
4. Shallower models perform better than deeper models when the data is less.
5. Accuracy of models trained using transfer learning outperformed models trained from scratch.

In [ ]: