Computer Vision and pattern recognition exam project

Eleonora Donadini

February 2020

1 Introduction

The aim of this project is to implement an image classifier based on convolutional neural networks trying each step to increase the accuracy of classification. The improvement are generated by the change of the parameters and approach adopted. The used dataset is the one provided by the assignment (from [Lazebnik et al., 2006]). It collects 15 categories and is already divided in training set and test set containing respectively 2985 imagines and 1500 imagines.

2 First task

The following model represent the blueprint to follow to build a first network.

#	type	size
1	Image Input	64×64×1 images
2	Convolution	8 3×3 convolutions with stride 1
3	ReLU	
4	Max Pooling	2×2 max pooling with stride 2
5	Convolution	16 3×3 convolutions with stride 1
6	ReLU	
7	Max Pooling	2×2 max pooling with stride 2
8	Convolution	32 3×3 convolutions with stride 1
9	ReLU	
10	Fully Connected	15
11	Softmax	softmax
12	Classification Output	crossentropyex

Figure 1: Layout of the provided CNN

2.1 Data preprocessing

As we can see the imput image is 64x64x1. The first two numbers refers to the size while the third to the channel (color or black and white) of the image. For that reason in order to feed the images to the network we have to resize the them to 64x64 and to set the channel to 1 (black and white image).
We must be careful to perform an anisotropic transformation rescaling the whole image independently along x and y to get the proper size, unless we could lose image data. Python provide proper functions to perform these tasks.

The foward step is to split the provided training set in 85% for actual training set and 15% to be used as validation set.

```
X_train, X_validation, y_train, y_validation = train_test_split(
    X_train_raw, y_train_one_hot, train_size=0.85, random_state=42)
```

2.2 The model

As request we employ the stochastic gradient descent with momentum optimization algorithm, using the default parameters, except for those specified. We se minibatches of size 32 and initial weights drawn from a Gaussian distribution with a mean of 0 and a standard deviation of 0.01. We also set the initial bias values to 0. Our final model is:

```
base_model = Sequential([
    Conv2D(8, 3,strides=1, padding='same',input_shape=(64,64,1)),
    Activation('relu'),
    MaxPooling2D(pool_size=2,strides=2),
    Conv2D(16, 3,strides=1, padding='same'),
    Activation('relu'),
    MaxPooling2D(pool_size=2,strides=2),
    Conv2D(32, 3,strides=1, padding='same'),
    Activation('relu'),
    Flatten(),
    Dense(15, activation='softmax',
        kernel_initializer=norm,
        bias_initializer='zeros')
])

base_model.compile(loss='categorical_crossentropy',
                   optimizer=sgd,
                   metrics = ['accuracy'])
```

2.3 Results

Here the plots of loss and accuracy over the epochs during the training, for both the training set and the validation set. The training stopped after 12 epochs. As we can see this model is probably affected by overfitting because the accuracy on the train set linearly increase while in the validation set stabilizes around 30%. This is probably due to fact that the data set uses for training is very small.

```
Train set dimintion:  1275 images
Validation set dimintion:  225 images
```

We reached the following results:

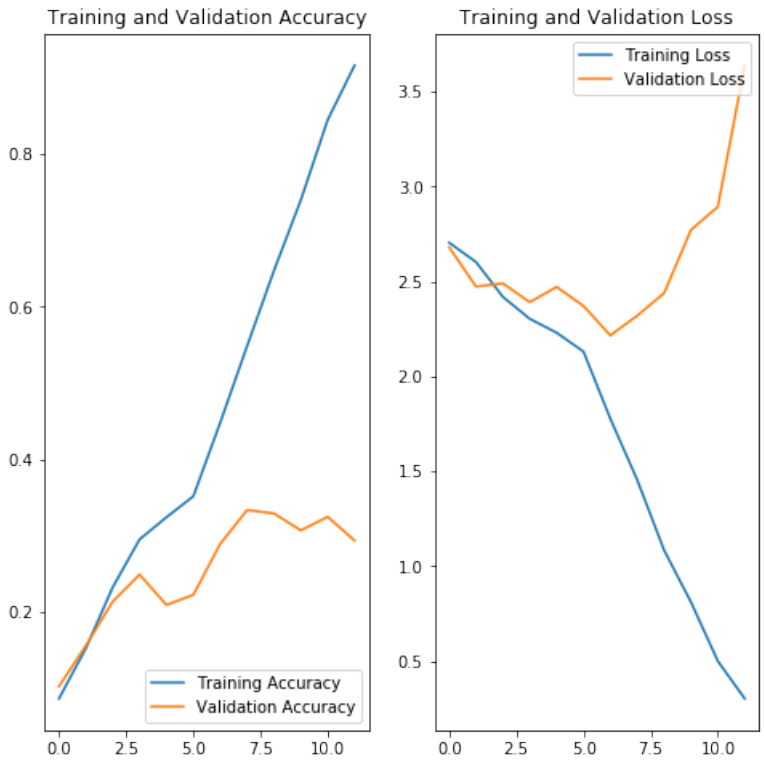

Figure 2: Layout of the provided CNN

In terms of test-loss and test accuracy, we reached the following results:

```
loss: 3.42
accuracy: 0.32
```

We can observe a very high loss. Due to the stochastic part of the model these results may vary but on average we can see that accuracy stabilizes around 30%.

We can now take a look to the confusion matrix.
We can observe that the accuracy can vary from class to class, for example, class " Highway" and " TallBuilding" have an accuracy of over 50%,it means that these two class are easily identifiable. Other classes, such as "Kitchen" or "Bedroom" have a very low accuracy, so there are much more uncertainty to identify.
Finally the following classification report summarize all of the score we got.

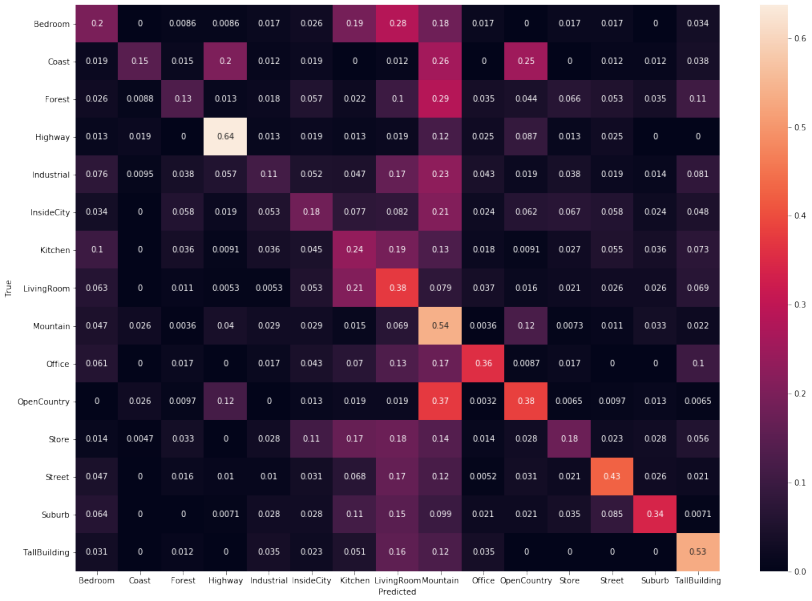

Figure 3: Layout of the provided CNN

Classification Report	precision	recall	f1−score	support
Bedroom	0.18	0.20	0.19	116
Coast	0.62	0.15	0.24	260
Forest	0.38	0.13	0.19	228
Highway	0.45	0.64	0.53	160
Industrial	0.29	0.11	0.16	211
InsideCity	0.26	0.18	0.22	208
Kitchen	0.12	0.24	0.16	110
LivingRoom	0.19	0.38	0.25	189
Mountain	0.22	0.54	0.31	274
Office	0.43	0.36	0.39	115
OpenCountry	0.42	0.38	0.40	310
Store	0.38	0.18	0.24	215
Street	0.54	0.43	0.48	192
Suburb	0.48	0.34	0.40	141
TallBuilding	0.53	0.53	0.53	256
accuracy			0.32	2985
macro avg	0.36	0.32	0.31	2985
weighted avg	0.38	0.32	0.32	2985

3 Second Task

The aim of this part is to exploit some more techniques in order to obtain a test accuracy of about 60%. We focus on:

1. data augmentation
2. batch normalization
3. change the size and/or the number of the convolutional filters
4. switch to Adam optimizer
5. dropout
6. ensemble of networks technique

3.1 Data augmentation

To reduce overfitting we increase the training data set applying the so called technique "data augmentation". With this technique we augmented the number of data with simple transformation: left-to-right reflections and rotation.

```
data_generator = ImageDataGenerator(
                        rotation_range=45,
                        horizontal_flip=True,)
```

Just with these simple modifications and using the same cnn developed in the first part, I reached a quite good implementation:

```
loss: 1.72
accuracy: 0.48
```

3.2 Batch normalization, Dropout and Adam Optimizer

Another two very common technique to improve results of a cnn are Dropout3 , which avoids the possibility of overfitting, and batch normalization, which controls the values of weights and bias avoiding the risk of gradient explosion. Moreover another thing to underline is the fact that I change the optimizer from sgd to adam with the default parameters.

3.3 Ensamble of Networks

Another very important technique to improve results is to train independently a several number of cnn of the same type (from 5 to 10 identical convolutional neural networks) and then find a method to "merge" different outputs in a smart way to make the results better and much robust.

```
Classification Report
              precision    recall  f1-score   support

     Bedroom       0.19      0.58      0.29       116
       Coast       0.30      0.89      0.44       260
      Forest       0.70      0.61      0.65       228
     Highway       0.73      0.59      0.65       160
  Industrial       0.47      0.32      0.38       211
  InsideCity       0.49      0.15      0.23       208
     Kitchen       0.20      0.18      0.19       110
  LivingRoom       0.35      0.41      0.38       189
    Mountain       0.64      0.34      0.44       274
      Office       0.63      0.57      0.60       115
 OpenCountry       0.66      0.36      0.46       310
       Store       0.61      0.42      0.50       215
      Street       0.88      0.55      0.68       192
      Suburb       0.75      0.76      0.76       141
TallBuilding       0.82      0.55      0.66       256

    accuracy                           0.48      2985
   macro avg       0.56      0.48      0.49      2985
weighted avg       0.58      0.48      0.49      2985
```

4 Third Task

In this last part I exploit the features of a pre-trained cnn in two different ways: 1. Freeze the weights of all the layers but the last fully connected layer and fine-tune the weights of the last layer with the same data as before. 2. Employ the pre-trained network as a feature extractor, accessing the activation of the last convolutional layer and train a multiclass linear SVM.

I decided to use a convolutional neural network called VGG16, which is included in keras package and it has got the following structure:

```
Model: "vgg16"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         (None, 224, 224, 3)       0
_________________________________________________________________
block1_conv1 (Conv2D)        (None, 224, 224, 64)      1792
_________________________________________________________________
block1_conv2 (Conv2D)        (None, 224, 224, 64)      36928
_________________________________________________________________
block1_pool (MaxPooling2D)   (None, 112, 112, 64)      0
_________________________________________________________________
block2_conv1 (Conv2D)        (None, 112, 112, 128)     73856
_________________________________________________________________
block2_conv2 (Conv2D)        (None, 112, 112, 128)     147584
_________________________________________________________________
block2_pool (MaxPooling2D)   (None, 56, 56, 128)       0
_________________________________________________________________
block3_conv1 (Conv2D)        (None, 56, 56, 256)       295168
_________________________________________________________________
block3_conv2 (Conv2D)        (None, 56, 56, 256)       590080
```

```
_________________________________________________________________
block3_conv3 (Conv2D)        (None, 56, 56, 256)       590080
_________________________________________________________________
block3_pool (MaxPooling2D)   (None, 28, 28, 256)       0
_________________________________________________________________
block4_conv1 (Conv2D)        (None, 28, 28, 512)       1180160
_________________________________________________________________
block4_conv2 (Conv2D)        (None, 28, 28, 512)       2359808
_________________________________________________________________
block4_conv3 (Conv2D)        (None, 28, 28, 512)       2359808
_________________________________________________________________
block4_pool (MaxPooling2D)   (None, 14, 14, 512)       0
_________________________________________________________________
block5_conv1 (Conv2D)        (None, 14, 14, 512)       2359808
_________________________________________________________________
block5_conv2 (Conv2D)        (None, 14, 14, 512)       2359808
_________________________________________________________________
block5_conv3 (Conv2D)        (None, 14, 14, 512)       2359808
_________________________________________________________________
block5_pool (MaxPooling2D)   (None, 7, 7, 512)         0
=================================================================
Total params: 14,714,688
Trainable params: 14,714,688
Non-trainable params: 0
_________________________________________________________________
```

4.1 Freezing the weighs and tuning

The first part is dedicated to train only the last fully connected component of the VGG16, while the remaining weights are freezed.
VGG16 has been written for resolving a classification problem between 1000 different classes, we have to change the last classification layer with a layer conform to our problem. Moreover the imput images have shape 224x224 so we adequate our image to the format the model requires. As loss function we decide to use categorical crossentropy.

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
vgg16 (Model)                (None, 7, 7, 512)         14714688
_________________________________________________________________
flatten_2 (Flatten)          (None, 25088)             0
_________________________________________________________________
dense_2 (Dense)              (None, 15)                376335
=================================================================
Total params: 15,091,023
Trainable params: 376,335
Non-trainable params: 14,714,688
```

We train the weights of the last fully connected layer for 20 epochs and using a batch-size equal to 32 and we obtain very good results:

```
loss: 0.48
```

```
accuracy: 0.83
```

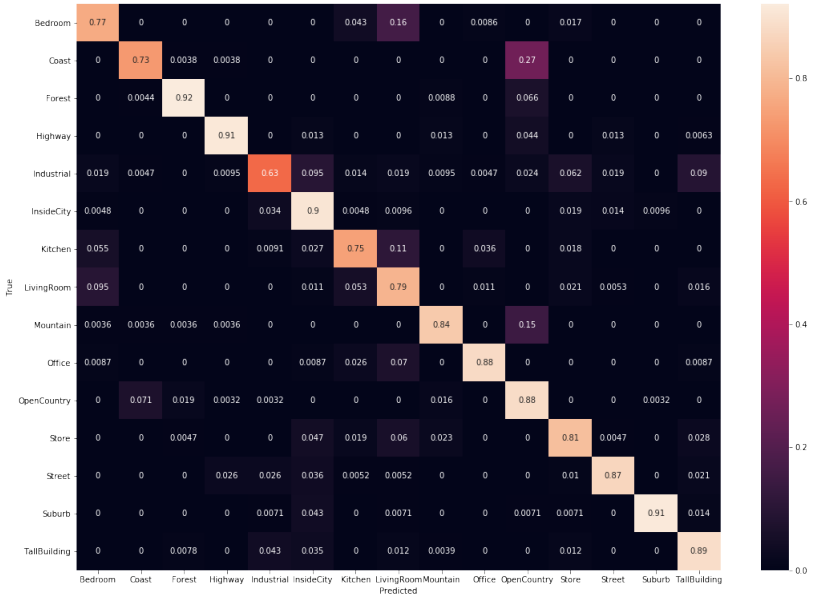

Figure 4: Layout of the provided CNN

4.2 Freezing the weighs

Last thing to try now is to exploit the cnn not for classifying directly classes of images, but to extract features to use after in another classification technique; in our case we will perform a linear SVM. In particular we will extract features obtained in the last convolutional layer of the VGG16. For each image, we will obtain 512 features, each of them is a 7x7 grid of values, so for each image we associate a feature represented by a 25088-dimension vector.
After that we simple use the features in a SVM. Being a multiclass classification problem, it is necessary to decide what strategy one has to use for classification; I personally adopted a one-against-rest approach which is implemented by the method fit of the library that we use.
We obtain very good results:

```
accuracy: 0.84
```

5 Conclusion

In this project, we will exploit all the most important feature of cnn for a multiclass classification problem and we have observed that techniques based

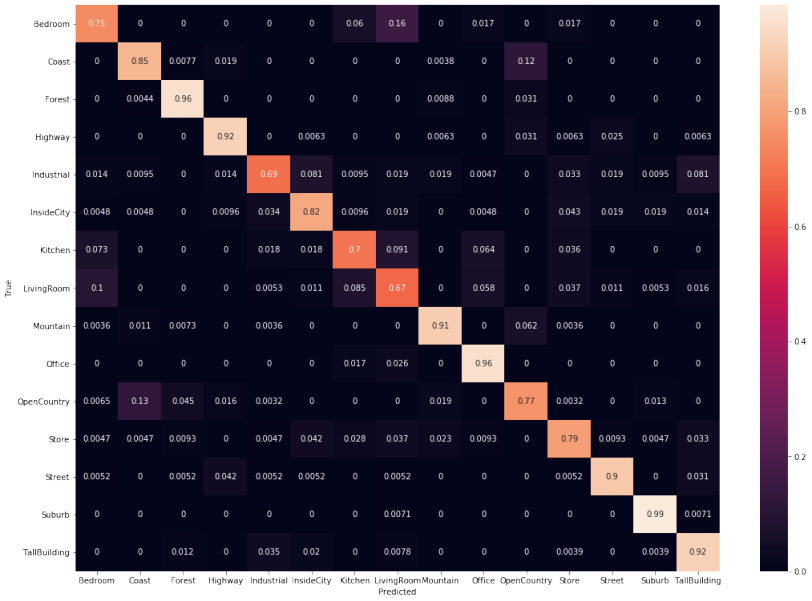

Figure 5: Layout of the provided CNN

on cnn are very powerful; A very shallow convolutional neural network was enough to obtain a quite good results and it has been possible improve these results with simple tricks like batch normalization or data augmentation. In the end, we applied transfer learning, in particular we extract features obtained by the last convolutional layer of VGG16 and we exploited them in a SVM, obtaining a very high test accuracy.