Ain Shams University
Faculty of Computer & Information Sciences
Scientific Computing Department

Neural Networks & Deep Learning
Sports Image Classification

By

SC_19

No	Name	ID
1	أحمد عبدالجواد رمضان عبدالجواد	20191700047
2	خالد شريف عبداللطيف عبد المجيد عزام	20191700225
3	سعد وليد سعد على ابوحسن	20191700284
4	عمر أحمد الشناوى عبدالحميد	20191700397
5	امير ممدوح حلمى شاكر	20191700138

Data preparation

1. Read image paths using the OS library.

2. Passed the paths to OpenCV to read images and resize them based on the Architecture we are going to use, (e.g (224, 224, 3) for ResNet, (32, 32, 1) for LeNet.

3. Created labels for CNN models, labels were one hot encoded (e.g [0, 0, 0, 1, 0, 0] for Swimming).

4. For Test images that were unlabeled, we extracted the image name in the data preparation process from the passed path to be used in the CSV result file.

5. Different Data Augmentation Techniques were used in the trials e.g (Random Horizontal and Vertical Flips, Random Rotation, and Random Crops . etc.)

6. We Tried Different Pre-Processing Techniques like normalization.

Models

We have conducted 28 Trials, in 5 different Architectures. ResNet was the architecture that achieved the highest testing accuracy **86%**.

1. **Lab Architecture**

   First Architecture we tried was the simple architecture provided in Lab 7. 3 Trials were conducted using this Architecture.

   The best Achieved Accuracy with this Architecture was 68% on the testing set.

   The Architecture was tested with and without normalizing the data. Data normalization didn't help in increasing the Accuracy.

   Architecture Summary:

   This Architecture consists of 5 Convolution Layers and 5 Pooling Layers, kernel sizes were (5, 5).

   Each Convolution layer is followed by a Pooling layer at the end 1 fully connected layer with Softmax as Activation Function.

2. **LeNet**

Second Architecture we tried was LeNet which is one of the first introduced Convolutional Neural Networks (1988).

3 Trials were Conducted using this Architecture.

Layer		Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	32x32	-	-	-
1	Convolution	6	28x28	5x5	1	tanh
2	Average Pooling	6	14x14	2x2	2	tanh
3	Convolution	16	10x10	5x5	1	tanh
4	Average Pooling	16	5x5	2x2	2	tanh
5	Convolution	120	1x1	5x5	1	tanh
6	FC	-	84	-	-	tanh
Output	FC	-	10	-	-	softmax

In our trials, we tried to increase the number of layers In the architecture following the same pattern, and increasing the image size, tried to use RGB instead of grayscale, and tried to change the activation function from tanh to ReLU.

The Best Achieved Accuracy yet was **62%** on the testing set.

3. **AlexNet (2012)**

The Third architecture we tried was AlexNet. 11 Trials were conducted with this architecture.

The architecture consists of 5 Convolution layers and 3 Max Pooling layers followed by a dropout by 0.5 and 3 fully connected layers.

Layer	# filters / neurons	Filter size	Stride	Padding	Size of feature map	Activation function
Input	-	-	-	-	227 x 227 x 3	-
Conv 1	96	11 x 11	4	-	55 x 55 x 96	ReLU
Max Pool 1	-	3 x 3	2	-	27 x 27 x 96	-
Conv 2	256	5 x 5	1	2	27 x 27 x 256	ReLU
Max Pool 2	-	3 x 3	2	-	13 x 13 x 256	-
Conv 3	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 4	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 5	256	3 x 3	1	1	13 x 13 x 256	ReLU
Max Pool 3	-	3 x 3	2	-	6 x 6 x 256	-
Dropout 1	rate = 0.5	-	-	-	6 x 6 x 256	-

With AlexNet we tried to augment the images we used three augmentation techniques Random Horizontal Flip, Random Vertical Flip, and Random Rotation. We tried to Randomly Crop images, but it didn’t help with the accuracy.

The suggested image size for AlexNet was 227 x 227 x 3, but when we reduced the image size to 128 x 128 x 3, we save training time and achieved better accuracy.

We even tried to do the same preprocessing which was done in the original paper that introduced AlexNet, but it didn’t increase the Accuracy that much.

We Tried to divide all the pixel values by 255 to normalize the data in the range between 0 and 1, it decreased the training time, but it only increased the accuracy slightly.

Our best Trial with AlexNet Achieved **75% Accuracy**, with these parameters:

- Image size (128 x 128 x 3)
- 30 epochs
- Random Horizontal Flip
- Random Vertical Flip
- Random Rotation Flip
- Adam Optimizer
- Early Stopping to prevent overfitting.

4. **VGG – 16**
   The Fourth architecture we tried was **VGG-16** which achieved our **2nd highest accuracy**. 3 trials were conducted using this architecture. The unique thing about this model is that it didn't focus on having a large number of hyperparameters but focused on having convolution layers of filters (3, 3) and max pool layer (2, 2), and followed this pattern throughout the architecture.

   With VGG- 16 we tried to load weights trained on the imageNet dataset and use transfer learning this increased the accuracy significantly. We used the same parameters which we used in AlexNet but with the original image size for VGG- 16 which is (224, 224, 3)

   The best Achieved Accuracy with this architecture is **81%**.

5. **ResNet**
   The Fifth architecture we tried was ResNet which achieved our **highest accuracy of 86%**. The unique about ResNet is that it allows training extremely deep learning models. And it introduced the concept of skipping connections. Which allows the high layers to learn as much as the lower layers.

   We used the concept of transfer learning and loaded weights trained on the imageNet set.

   In the highest achieved trial, we trained the model with 100 epochs, Adam optimizer, augmentation (Random flips and rotations), and early stopping to prevent overfitting.

Conclusion

- **From our trials, we found out that ResNet architecture is the most suitable for our task which achieved 86% accuracy on Kaggle.**
- **For some architectures reducing the image sizes significantly increases the accuracy.**
- **Right augmentations definitely increase the accuracy with high portions we found out that (random flips and rotations) are most suitable for our task.**
- **Data normalization reduces the training time a lot but we didn't get much increase in the accuracy out of it.**