

Dataset Preprocessing

- A good dataset is a key for successful model training.
 - This assignment guides you how to create a labeled dataset, which would later be used for supervised learning.
-

Overview

Image augmentations are used to increase the variability and robustness of a dataset

Goals - Create a larger dataset from a smaller one

Goals

- The original MNIST dataset includes images of single digits, in ten classes.
 - Your goal is to create a new three-digit image dataset, with 101 classes, based on the original MNIST dataset.
-

Tasks

1 - MNIST

- Download the mnist data using the code snippet below

```
import torchvision
mnist_data = torchvision.datasets.MNIST(root='',
download=True)
this downloads the data to root ('') folder. Change the
root to a convenient location.
```
- Show at least 5 samples of each class (0-9). Always add proper titles and choose the correct color map
- Your new dataset should include the three-digits numbers: [000, 001, ..., 055, ..., 099, 100], forming 101 classes.
- Each class should have at least 4000 different samples.
- Use the augmentations you created on lab 3 (Flip, Rotate, Gaussian Noise (adding to the image as noise), Gaussian-Blur (convolution with a gaussian filter), Median-Blur and Zoom) on your new dataset and show some results and explain which augmentations are applicable to this specific dataset, which are not and why. For example rotation by 180 degrees is problematic because it turns 6 into 9.
- you should add about 10 new augmentations (you may use OpenCV / Albumentation). The total number of augmentations will be 15.
- Display typical results of your dataset, demonstrating the variability of the new dataset.
- Write a Dataset class (see reference in LAB 4) for the MNIST101 Dataset you implemented. Creating at least 4000 samples for every class should be done prior to the implementation of the dataset class.
- You should run your code on mnist10 train subset for the train subset of mnist101 and on the test subset of mnist10 for the testset of mnist101.
- Write a detailed summary of your work.

3 - Summary

- Write a short summary of everything you've done in this work