Neural Networks for Images - Exercise #3

submission date: 23/6/2021

Programming Task: Deep Image Priors

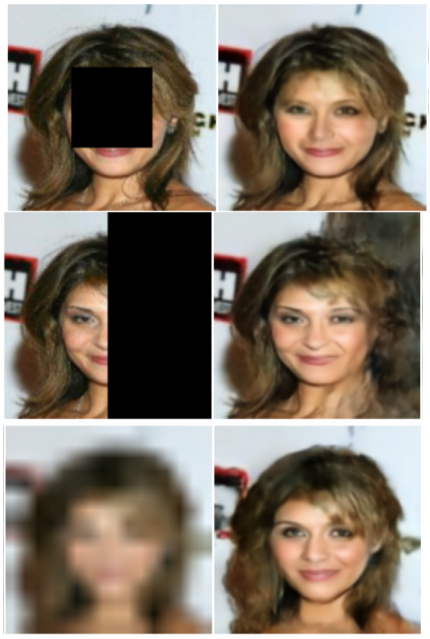

Image priors offer the missing tool for solving every image restoration task, as long as its likelihood (data) term can be formulated. As shown on the right, this includes in-painting and deblurring. In this exercise you will construct such a prior using an auto-encoding scheme, similar to the bijective image prior (BIG) we saw in class, but without the GAN loss. You will need to think of a suitable replacement for the latter.

As in the previous exercises, you will implement this model, evaluate it and document the results along with your explanations and design considerations in a pdf report that you will submit (see Submission Guidelines below).

Specifically we will follow the first steps of the BIG derivations, namely, we'll define the prior $P(I) \sim \exp(-|| D(E(I)) - I ||/T)$, where $E(I)$ is an encoder, $D(.)$ is a decoder, and T is some meta-parameter that controls the sensitivity of the prior to the reconstruction error. In order for the prior to function, it needs to assign low density values to outlier images, meaning that $D(E(.))$ should not be able to properly reconstruct them, and vice versa, inliers should be accurately reconstructed.

By training $D(E(.))$ as an auto-encoder, i.e., by minimizing its reconstruction loss over a target class of images, we'll ensure one requirement but not the other, i.e., it may still accurately reconstruct some out-of-class images. The BIG construction ensures that all $D(.)$ can only span the target set of images using adversarial training and ensures the encoder $E(.)$ maps only into $D(.)$'s latent set. We'd like to use this construction, i.e. minimize the reconstruction loss over a bounded set $[0,1]^d$ in latent space, but avoid using the costly GAN training to control $D(.)$'s image (the manifold it spans).

In order to come up with an alternative solution we will start by investigating what actually happens when training this auto-encoding problem through this bounded set, i.e.,

1. **Naive training.** You will train D and E to minimize $|| D(E(I))-I ||$ over the target class of images you'll use - which will be MNIST - and a sigmoid at the end of the encoder network E.
2. **Analyze the Data.** Use a 2D scatter plot of the encoded images in several planes in latent space. The latent space dimension d should be between 8 and 12, you can pick several pairs of coordinates $0 <= i,j < d$ in this space and plot the latent codes of say 300 images.

3. **Define Failure Scenarios.** Explain how the points are distributed in [0,1]^d - and why this can allow outliers to be properly reconstructed.
4. **Find a Solution.** Come up with a simple analytical loss that will fight the shortcoming observed in 3 and train the networks with respect to this loss term combined with the reconstruction loss.
5. **Image Restoration.** Use both the naive and your networks (from items 1 and 4) to solve the two following restoration problems:
    a. *Image denoising:* add i.i.d Gaussian noise with N(0,s) with some small standard-deviation s, and use Bayes Law to formulate the posterior distribution and minimize its -log
    b. *Image in-painting:* do not add noise to the image in the likelihood term, but use a zero-one mask to "shut-down" the data over an interesting set of pixels.

Report all the results obtained in these items (show the images and explain which network performed best).

We expect you to report and elaborate on every practical task in the report, §using your own words and your own analysis of what you've done. Include everything that you think is crucial for us to understand your way of thinking.

Theoretical Questions:

1. Explain why Gram matrices are suitable for capturing the characteristics of a texture. What advantage is gained by computing the Gram matrices for multiple convolution layers? What would happen if we use only the fine resolution layers? What would happen if we only use the deeper layers?
2. When doing style transfer using AdaIN, how is a texture characterized? Seemingly, this characterization ignores dependencies between different feature channels, discuss whether this is really so.
3. Non-adversarial generators:
    a. Which of the methods we discussed in class provides the sample's density value, i.e. P(I) where I is the generated image? Explain.
    b. Which methods can, in principle, improve and improve their accuracy as their network size and the number of training examples get higher, and which do not improve. Explain.
    c. Which activation function is suitable for the GLOW method: ReLU or LeakyReLU? Why?
    d. The IMLE method is typically implemented using an approximate nearest neighbour (ANN) search instead of the summation over all the points in the density estimation formula. ANNs are known to fail at high dimensional spaces. How would you suggest implementing the IMLE using ANN?
4. Assume that we would like to extend the Deep Image Prior (DIP) idea to general signals (not images) where the generator is simply a multi-layered perceptron (network of FC layers). Do you expect this approach, with such a network to provide regularization of some form? explain.

Submission Guidelines:

The submission is in **pairs.** Please submit a single zip file named “ex3_ID1_ID2.zip”. This file should contain your code, along with an ”ex3.pdf” file which should contain your answers to the theoretical part as well as the figures and analysis for the practical part. In addition, please include in this compressed archive a README with your names and cse usernames.
Please write readable code, with documentation where needed, as the code will also be checked manually.