Implicit Neural Representation with periodic activation function

Francesca Pistilli (francesca.pistilli@polito.it)

OVERVIEW

An interesting application of neural networks that has been recently investigated is to learn continuous, differentiable and memory-efficient signals representations, also called implicit neural representation. The goal is to learn a neural network able to parametrize a signal, i.e. it takes as input spatial or spatio-temporal coordinates and provides as output the desired signal. Architectures based on multilayer perceptrons with ReLU activation functions typically fail to represent details and have some intrinsic limitation. Recently, an innovative alternative based on periodic activation functions has been proposed. In this project you will focus on the architecture with sinusoidal non-linearities and images as target signals. First of all you will familiarize with the network and experiment it in well-established tasks, then you will adapt and extend it to some novel applications.

GOALS

1. Study and investigate the topic of implicit neural representation.
2. Read and understand the article SIREN that presents the network with periodic activation functions.
3. Implement a basic version of the SIREN network for some simple applications and try to emulate the results in SIREN.
4. Implement a SIREN-based network to solve one of the two proposed tasks: Single Image Super Resolution or Video Frame Interpolation

REQUIREMENTS

1. Implement a basic 2 hidden layers SIREN in PyTorch. A SIREN is just a fully connected network that exploits a sine function as non-linearities, but be careful to weight initializations and sine frequencies. You can find all the detailed information in SIREN.

2. Train and test your classic SIREN for two basic tasks, such image fitting and poisson image reconstruction, and try to replicate the results in SIREN. I suggest you tasks that involve images but if you are curious feel free to experiment also with other types of signals, in SIREN you can find many alternatives.
3. Compare the obtained results with a fully connected network with ReLU as activation function. Make sure to have a fair comparison, so same number of parameters and layers.
4. Make three ablation studies changing the weights initialization of the first layer, of the hidden layers and the ω_0 parameter (the ω_0 utilization is well explained in the paper, this parameter determines the frequency of the sine). The aim of these experiments is to well understand the important features of the network.
5. Implement a SIREN-based network to address the task of Single Image Super Resolution (SISR) or Video Frame Interpolation (VFI). A detailed explanation of the two tasks is reported in the PROJECT section.

Tips:

- The skimage and skvideo libraries are useful to directly load some well-known images (such as the Cameraman) or videos.

What to submit:

- Working pytorch scripts for all the required steps.
- A complete pdf report. The report must contain: a brief introduction, a related works section, a methodological section, an experimental section with all the results (for all the required steps) and a discussion. End the report with a brief conclusion.

PROJECTS

Single Image Super Resolution

The single image super resolution task consists in obtaining a high-resolution (HR) image having as information the low resolution version (LR). Typically the low resolution consists in an image of smaller size or with some inserted degradation (e.g. blurring). Your goal is to implement a SIREN-like network, able to provide a continuous high resolution representation of an image. You can choose one of the well known images as a target (Lena, Cameraman etc).

Remember that in your setting the coordinates of a 2D grid are the input of the network, not the image!

The idea to implement this task is to train the network to represent a low resolution image with a corresponding low-resolution coordinates grid as input, and test it with a denser grid in order to obtain the high-resolution image.

Report the results in terms of PSNR and make comparisons with one of the state-of-the-art methods reported in Review SISR.

Video frame interpolation

A video is a collection of subsequent images. The video frame interpolation task consists in predicting an intermediate image having two subsequent images.

Your goal is to find an implementation of the problem suitable with the SIREN network.

Remember that in your setting the spatio-temporal coordinates of a 3D grid are the input of the network, not the video!

The idea is to train the network to learn a continuous representation of the video having two images to constrain the optimization procedure. Then, during the test phase, you extrapolate a frame that is temporal intermediate between the two originals.

Report the results in terms of interpolation error and make comparisons with one state-of-the-art method, you can look at the review reported in Review VFM.

BIBLIOGRAPHY

[SIREN] "Implicit Neural Representations with Periodic Activation Functions", Vincent Sitzmann et al.

Also take a look at the website!

[ReLU PE] " Nerf: Representing scenes as neural radiance fields for view synthesis", Ben Mildenhall et al.

[Review SISR] "Deep Learning for Single Image Super-Resolution: A Brief Review ", Wenming Yang et al.

[Review VFI] " Review: AdaConv — Video Frame Interpolation via Adaptive Convolution (Video Frame Interpolation) ", Sik-Ho tsang .