CMSC828I Project Proposal - Meta Learning Feature Based INR

Darshit Miteshkumar Desai
darshit@umd.edu

Hitesh Kyatham
hkyatham@umd.edu

Suriya Suresh
jsuriya@umd.edu

Vinay Krishna Bukka
vinay06@umd.edu

Abstract

The document describes the project proposal on the topic of "Meta-learning for feature grid based Implicit Neural Representations (INRs)". The paper describes a brief introduction on the state of the field by mentioning its applications in 3D computer vision, Image and Video compression and also enumerates INR and it's disadvantages and proceeds on to explain improvements done on top of that. The project intends to combine the various improvements done on INR to form a new algorithm which combines Meta-learning, which is a task of predicting weights of a neural network and feature grid based INRs. Essentially our objective is to use Meta Learning to predict weights of a Multi Layer Perceptron which uses Feature grid based inputs rather then the conventional coordinate based inputs.

1. Introduction and Motivation

In recent years, Neural Implicit Representations have been used and proposed in various research work as a way of representing various modalities of data. Implicit Neural representations (INR) is a mapping of input to output using continuous implicit function which is learnt during the training process. This type of implicit function is used a lot in 3D computer vision as a representation apart from point clouds, voxels and surface meshes. The advantage of this representation in the context of 3D vision is that the memory size is significantly reduced and the representation at the same time is more denser and visually appealing. For the case of memory, Representations like Voxels occupy cubic memory for storing the 3D object, while representations like point clouds are sparse and representations like surface mesh are usually difficult to obtain as they are usually rendered from point clouds or other intermediate representations which is memory and time intensive. INR gives a direct method of representing all these representations in the form of probabilities of the points sampled in 3D space. The same approach can be replicated for other modalities of data, like Images and Videos. In images, a dedicated INR can be trained to predict and map R^2 space (x, y) coordinates to (R, G, B) space in R^3. In case of videos spatio-temporal data is encoded (i.e., (x, y, t)) to a continuous learn-able implicit function which predicts video frames based on it. This in a way acts as a compression algorithm where Gigapixel sized images and other large sized videos can be represented as a light weight neural network.

The disadvantage of the above light weight Multi layer Perceptrons (MLP) are that if the same compression task is to be scaled for dataset level images and videos, the same would require dedicated MLPs for all images or videos in the dataset. Many papers come up with some pruning techniques to reduce the number of neurons/weights required to reduce the size of MLPs but there is an additional amount of training required to fine tune the pruned INRs. To overcome this disadvantage, the concept of Meta-learning is used. Meta-learning is a task of predicting weights of a neural network. Specifically in case of INR, a dataset of images is passed as an input of a learning based algorithm to predict weights of a MLP which renders an image, video or 3D data. Although Meta learning helps in pruning and finetuning the INR models by predicting weights which apply to a dataset of images, but inherently computationally intensive since the algorithm has to take into consideration each pixel to reconstruct an image. This is especially cumbersome in dataset operations where a dataset which has a lot of matching features like dataset of cars might not need pixel wise sampling.

To overcome the disadvantage of high amount of computations for reconstruction, a feature grid based INR is proposed where during training a set of grid features representing the dataset is trained and modified. The feature grid represents a large portion of a subset of pixels in case of images and if those pixels don't change much in the given grid the features are not modified much. The input to this type of networks is a feature grid and the output is an image (or any other data modality). In this scenario although the computations are reduced there is an increase in the memory required to store the feature grids along with the network weights.

The most direct solution to the increase in memory because of model weights is to use Meta learning for learning a network which predicts weights in real time during test for the feature grid based INR. This novel combination of meta-learning and feature grid based INR is what would be explored in this project.

2. Approach

2.1. Implicit neural representaions.

(INR) Recent work on implicit representations for shape parts, objects or scenes has shown the power of these networks to represent the inputs to these networks as a function to map some quantity of intereset defined by F as mentioned in [9].

We first plan on implementing INR and overfit it to one single image to improve our understanding of INR and there by deepen our understanding of INR. Where in we take some image as ground truth and try to overfit our model on to represent this image.

2.2. Meta-Learning with INRs

INRs are memory intensive and when using it in approaches such as the image representation, the memory taken up by the INRs(weights) is comparatively more than the actual memory of the image when stored in standard JPEG format. This is just one such instance, but overall when using INRs for huge datasets, it becomes intensive with regards to both memory and computations needed. To alleviate the memory issue of INRs when applied to datasets, we can use the concept of meta-learning with INR, where in the weights of the INR can be predicted using some means such as Hypernetworks. These weights can be used in the INR directly and it is noted that the INR with such weights can be used with bare minimum training. We plan to implement Meta-Learning with INR to understand how the memory issues can be resolved and to understand more about meta learning implementation.

2.3. Feature-grid with INRs

As we know that the INR are computation intensive. Feature-grid INRs which typically map a MLP $g(.)$ with parameters ϕ, to represent a given input x [4]. Feature grid provide an advantage as they club the additional trainable parameters as a grids where each gradient propagated backwards through the network, only a very small number of grids need to be updated for each back-propagated to the encoding. This way, although the total number of parameters is much higher, the number of FLOPs and memory accesses required for the update during the training can be reduced significantly as mentioned in [6]. So the overall size of MLP can be reduced there by reducing the overall computations required in case of large data sets. This approach can be used when training a huge datasets. As our next step we plan on implementing INR with Feature-grids so that we can understand the importance of feature grids and how this implemetation can alleviate number of computations needed in an INR model.

Finally, we plan to combine both the approaches of Meta-learning and Feature-grids with INR to counter both the issues of INR without loosing the quality of the representations. Here we plan to first train some hypernetworks to predict some feature-grids which will then be used in an INR to represent the inputs.

3. Existing Work Done

Our project proposal aims to combine meta-learning with feature-grid based INRs to improve the slower training speeds of INRs while maintaining the advantages that these pose over coordinate-based INRs.
We intend to start from coordinate-based INRS, referring upon existing implementations of state of the art INRs like SIREN [9] and NeRF[8].Afterwards scaling up to feature grid based INR implementations like SHACIRA [4] and Zip-NeRF [1] to further our understanding of such INRs.
Girish *et al*. [4] proposed the use of learn-able multi-resolution feature grids.These feature grids store features at various Levels-of-Details which are then concatenated into a MLP to construct a signal.These have proven to have fast training times and tend to require massive memory for storing the feature maps.
There are existing works that have done the concept of meta-learning for INRs such as MetaSpareINR [5] and TransINR[2]. They have been able to build upon the faster performance of hyper-networks in training and combined it with practical benefits of INRs such lower data requirements, scalability to high resolutions and ability to parameterize input as continous functions.
Jacho *et al*. [5] uses the meta-learning framework MAML [3] which is general and model agnostic which aims to make tuning of the trained model easy. MAML is used to learn the INR that can fit a signal efficiently and then pruning is done to remove connections of layers that have the smallest weights and this step is repeated for fixed number of iterations till global sparsity level is reached.
Chen *et al*. [2] proposes using transformers for gradient based meta-learning of all weights for an INR.The transformer takes in images as in form of data tokens and init tokens representing the weights of the INR which builds seeks to transfer observations of interaction of data and init tokens to weights of INR and get mapped into the INR weights with correct dimensionality.This performs better compared to other gradient based meta-learning methosd such as MAML [3] and Reptile [7].

4. Achievable and Ambitious Goals

4.1. Achievable plan

In the achievable plan, we are going to follow the step-wise method mentioned in the approach where our intermediate milestones will act as learning milestones for the group members and build on each other as modules to be integrated at the final stage, We plan to test and train on the image compression modality of the network. The step-wise plan of implementation includes training and testing a simple Implicit Neural Representations for images, which would then be combined with a meta learning based weight prediction model. Upon successfully doing both the above things, a new module for feature grid based Implicit neural representations would be built as a learning milestone for grasping the intricate details of implementing such a network. At the end, the same would be combined with meta learning model for training and testing on images.

4.2. Ambitious Goals

Depending the success of our achievable goals we propose the following goals for improving upon our base goals:

- Fine-tuning the parameters of the hyper-network to produce faster training times and improve the estimated weights for the INR.
- Optimizing the pipeline to reduce the memory and computational resources
- Enabling the ability to use other source of input data such as videos and 3D point clouds.

5. Evaluation Metrics and Datasets

First, through a basic implementation of INR, the model built will be tested on datasets such as CelebA for images, UVG dataset for Videos and evaluate quantitative metrics such as PSNR (Peak Signal-to-Noise Ratio) based on [9]. The adapation of meta-learning based on the simple INR built will be tested on CelebA, Imagenet dataset and calculate the PSNR metric on different weight groups as in [2]. The grid based INR is planned to be tested on different metrics such as PSNR, Perceptual similarity, SSIM with reference to [4]. Finally, after the each step of implementation when the model representing the combined pipeline of Meta-Learning with Feature Grid based INR is built, it is planned to be tested on CelebA and Imagenet dataset to be tested on different images and calculate different quantitative metrics such as PSNR and compared with the results of individual implementation metrics as mentioned above. As mentioned in the ambitious goal, the same datasets and some datasets for videos such as KODAK are planned to be tested with the model after achieving final fine tuned parameters.

References

[1] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Zip-nerf: Anti-aliased grid-based neural radiance fields. *ICCV*, 2023. 2

[2] Yinbo Chen and Xiaolong Wang. Transformers as meta-learners for implicit neural representations, 2022. 2, 3

[3] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. *CoRR*, abs/1703.03400, 2017. 2

[4] Sharath Girish, Abhinav Shrivastava, and Kamal Gupta. Shacira: Scalable hash-grid compression for implicit neural representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 17513–17524, 2023. 2, 3

[5] Jaeho Lee, Jihoon Tack, Namhoon Lee, and Jinwoo Shin. Meta-learning sparse implicit neural representations. *CoRR*, abs/2110.14678, 2021. 2

[6] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *CoRR*, abs/2201.05989, 2022. 2

[7] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms, 2018. 2

[8] Liyue Shen, John Pauly, and Lei Xing. Nerp: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–13, 2022. 2

[9] SIREN. https://www.vincentsitzmann.com/siren. 2, 3