**On Few-Shot Learning with Neural Processes and Meta Learning**

* **Anirudh Suresh, Srivatsan Srinivasan**

[**anirudh\_suresh@college.harvard.edu**](mailto:anirudh_suresh@college.harvard.edu)**, srivatsansrinivasan@g.harvard.edu**

**INTRODUCTION**

TraditionalDeep Neural Networks have been inefficient when trained on small datasets. The search for more data-efficient learning in neural networks has carved out two kinds of approaches:

* Meta Learning - A two-phase learning process where the first phase learns the common abstract statistics of a generic domain and the second phase learns a function for a specific task with a small number of examples and exploiting the domain-wide statistics already learned. Standard examples include MAML and few-shot autoregressive density estimations [1,12].
* Probabilistic Learning - Specifying a distribution over functions (i.e. stochastic processes) where a practitioner’s prior knowledge is captured via the distributional assumptions on the function prior. GPs [13] are classic examples of this approach, but GPs become computationally intractable when the posterior inference is performed on a large dataset. Neural Processes (NPs [5], CNPs [6] followed by ANPs [4]) are a family of models that represent solutions to the supervised problem, and an end-to-end training approach to learning them, that combine neural networks with features reminiscent of Gaussian Processes.

In this project, we intend to explore a couple of state-of-the art works in neural processes [4] and unsupervised task learning+MAML [3] and add a few incremental improvements (see next section) to these works, study the efficacy and weaknesses of these models for few-shot learning tasks by running new experiments on text datasets which have structurally different properties compared to the relatively more structured image datasets (focus of original works).

**PROJECT CONTRIBUTIONS**

Our contributions in this project can be summarized as follows. In this section, we are not committing to all these proposed ideas and would love to cover as many as possible within the 40-day time period we have for this project. (our use of the \* symbol indicates our priority, *with more \*s indicating higher priority* - we would love any instructor feedback on re-prioritizing these tasks as well.).

* ***Model 1 : More Expressive Attentive Neural Processes(ANP)*** - We aim to add the following enhancements over the original ANP - a.) \*\*\*\* Build self-attention in the decoder of ANPs to increase their expressiveness, drawing inspiration from the text [7,8] and image transformers [16], b.) \*\*\*\* build cross-attention into the latent path and model the dependencies across the resulting local latents along with a global latent variable learns the overall structural dependencies, similar to the Neural Statistician setting [15].
* ***Model 2 : Unsupervised Task Learning + MAML -*** We aim to replicate the work of unsupervised task learning for MAML [3] and iterate upon it (as time permits) to a.) **\*\*\*** incorporate other advances in unsupervised learning (clustering, representation learning etc.) to the first phase of unsupervised task learning from raw data, b.) **\*\*** In the original paper, the task learning and meta learning are trained one after the other - we would want to jointly train end-to-end the task learning (including soft clustering with differentiable models such as GMMs) and meta-learning (with MAML) phases. The aim of the latter is to learn a task representation that benefits downstream meta-learning and understand the trade-offs between the two approaches. ***As this idea is relatively less concrete, we would like to receive more instructor feedback on the feasibility, potential advantages or pitfalls over the original paper.***
* ***Baselines, Study of pros and cons with experiments on Images (Continuous and relatively more structured) vs Text (Discrete and less structured)*** :
  1. **\*\*\*\*** First of all, we intend to test all our claims on the 1-D synthetic regression data as is done in other papers to compare and validate baseline performance. For our baselines, we intend to recreate or re-use most of the data+code from the corresponding works to ensure that the results are trustworthy.
  2. **\*\*\*** After the baseline experiments, we intend to focus on testing and comparing these two classes of models for few-shot learning on continuous and relatively more structured data like images vs. discrete and less structured data like text - All the results in the existing Neural Processes literature have been on images and image completion which inherently are more structured than text data. We are planning to test these models on language tasks such as sentence completion, multi-domain sentence classification etc. to see their strengths and weaknesses in a variety of language-related tasks with more complicated structures and function posteriors. With regard to datasets, for images we can continue using the celeb-A face-completion and MNIST datasets for regression; OmniGlot for classification; meanwhile, for text, we intend to use one of Penn TreeBank or SNLI (Stanford Natural Language Inference) datasets for sentence completion tasks and IMDB, Yelp Reviews datasets, etc. for classification.
     + The aim is to break down the existing plus new components we are adding to both these models and study their individual contributions to the final learning outcomes, addressing areas where they are expected to fail and suggest (and if possible, implement) ideas to address these discovered weaknesses. Combining these incremental changes and performing this study is expected to throw open new research directions for us to pursue in the future.

**RELATED WORK** (to keep this paper short, we will add a more detailed version in the final paper)

**MAML and unsupervised task learning -** The base model-agnostic meta-learning algorithm was first proposed by Finn et al [1] and discusses a training schema that develops a meta-learner which excels in few-shot learning for related new tasks. Manually specifying a good task set for meta learning is hard in several practical applications and Hsu et al (2018) [3] presents an unsupervised method of identifying and constructing tasks from a given dataset using techniques from unsupervised representation learning leveraging the works of [2,10,11,12].

**ANP and Transformers -** Attentive neural processes were first introduced in Kim et al (2018) [4] as an improvement over conventional NPs [5,6]. They leverage self-attention within context observations and cross-attention between context and targets to effectively learn target query-specific representations from training data. Notably, ANPs combine the latent (no-cross attention with target queries) and deterministic paths through a vanilla MLP. To increase the expressiveness of the network, we intend to implement self-attention in the decoder and cross-attention in the latent path, similar to the work of Lee et al (2019) [8] - the set transformer that offers a permutation-invariant encoder-decoder architecture, with attention based encoders and decoders, an idea largely influenced by Vaswani et al (2017) [7]. Similarly, image generation using only attention-based transformers has been proven to be successful in Parmer et.al. 2018 [16].

**Bibliography (with informal citations - we will include formal citations in the final paper)**

1. Finn, Chelsea et al, 2017. “Model-Agnostic Meta Learning for Fast Adaptation of Deep Networks.”
2. Caron, Mathilde et al, 2018. “Deep Clustering for Unsupervised Learning of Visual Features.”
3. Hsu, Kyle et al, 2018. “Unsupervised Learning via Meta-Learning.”
4. Kim, Hyunjik et al 2018. “Attentive Neural Processes.”
5. Garnelo, Marta et al 2018. “Neural Processes.”
6. Garnelo et.al. 2018. “Conditional Neural Processes”.
7. Vaswani, Ashish et al 2017. “Attention Is All You Need.”
8. Lee, Juho et al 2019. “Set Transformer: A framework for Attention-based Permutation-Invariant Neural Networks.”
9. Berthelot, David et al 2018. “Understanding and Improving Interpolation in Autoencoders via an Adversarial Regularizer.”
10. Donahue, Jeff et al, 2017. “Adversarial feature learning.”
11. Dumoulin, Vincent et al, 2017. “Adversarially learned inference.”
12. Chen, Xi et al, 2016. “InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets.”

# Scott Reed et.al. 2017. “Few-shot Autoregressive Density Estimation: Towards Learning to Learn Distributions”

1. Rasmussen, C. E. and Williams, C. K. Gaussian processes in machine learning. In Advanced lectures on machine learning, pp. 63–71. Springer, 2004.

# Harrison Edwards, Amos Storkey. “Towards a Neural Statistician”

1. Niki Parmar, Ashish Vaswani et.al. “ Image transformer” In ICML, 2018.