

# A short story on multi-task learning and meta-learning

***Multi-Tasking Learning** is one of the supervised learning techniques that involve fitting a model on one dataset that addresses multiple related problems. This provides an opportunity to tackle many of the challenges simultaneously including data and computation bottlenecks. Our discussion is structured according to a partition of the existing deep MTL techniques into three groups: architectures, optimization methods, and task relationship learning*

## Architectures:

1. Partitioning the network into task-specific and shared components in a way that allows for generalization through sharing and information flow between tasks, while minimizing negative transfer.
2. **Shared Trunk:** The network contains multiple shared trunks, and each task-specific output head receives as input a linear combination of the outputs of each shared trunk. The weights of the linear combination are computed by a separate gating function, which performs a linear transformation on the network input to compute the linear combination weights. The gating function can either be shared between all tasks so that each task-specific output head receives the same input, or task-specific so that each output head receives a different mixture of the shared trunk outputs
3. **Cross-Talk:** A Cross-Stitch network is composed of individual networks for each task, but the input to each layer is a linear combination of the outputs of the previous layer from every task network. The weights of each linear combination are learned and task-specific so that each layer can choose which tasks to leverage information from.
4. **Prediction Distillation:** In an MTL setup for jointly learning depth prediction and semantic segmentation, discontinuities in the depth map imply likely discontinuities in semantic segmentation labels, and vice versa.
5. **Task Routing:** Despite their success, shared trunk and cross-talk architectures are somewhat rigid in their parameter sharing scheme. A fine-grained parameter sharing between tasks that occurs at the feature level instead of the layer level. The novel component of this architecture is the Task Routing Layer which applies a

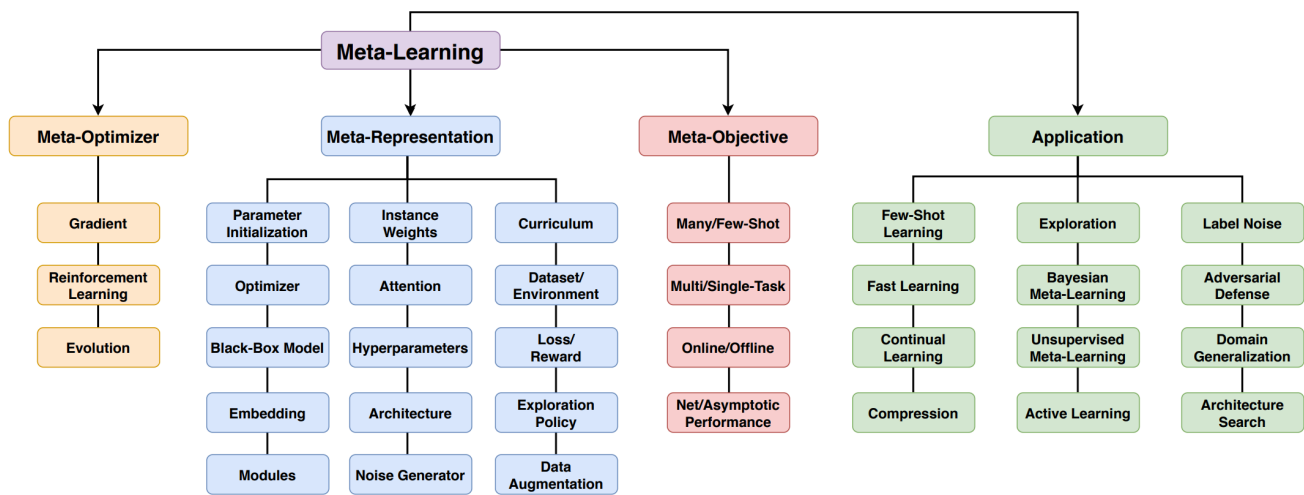
task-specific binary mask to the output of a convolutional layer to which it is applied, zeroing out a subset of the computed features and effectively assigning a subnetwork to each task which overlaps with that of other tasks.

## Optimization for Multitask learning

1. **Loss Weighting:** A very common approach to ease multi-task optimization is to balance the individual loss functions for different tasks. When a model is to be trained on more than one task, the various task-specific loss functions must be combined into a single aggregated loss function which the model is trained to minimize.
2. **Regularization:** Regularization has long played an important role in multi-task learning, mostly in the form of soft parameter sharing.
3. **Task Scheduling:** Task scheduling is the process of choosing which task or tasks to train on at each training step.
4. **Gradient Modulation**
5. **knowledge distillation:** instill a single multi-task “student” network with the knowledge of many individual single-task “teacher” networks.
6. **Multi-Objective Optimization.**

**Meta-Learning** in other terms *Learning to learn*, an automatic learning where there is no intervention need in the tasks like tweaking the model, choosing params. Meta-learning provides an alternative paradigm where a machine learning model gains experience over multiple learning episodes — often covering a distribution of related tasks — and uses this experience to improve its future learning performance.

6



## Methods/Architectures

1. Transfer Learning can refer to a problem area, meta-learning refers to a methodology that can be used to improve TL as well as other problems. TL as a methodology is differentiated to meta-learning as the prior is extracted by vanilla learning on the source task without the use of a meta-objective. In meta-learning, the corresponding prior would be defined by an outer optimization that evaluates how well the prior performs when helping to learn a new task, as illustrated, e.g., by MAML
2. Domain-shift refers to the situation where the source and target tasks have the same classes but the input distribution of the target task is shifted concerning the source task leading to reduced model performance upon transfer. DA is a variant of transfer learning that attempts to alleviate this issue by adapting the source-trained model using sparse or unlabeled data from the target.

3. **Continual learning (CL)** Continual and lifelong learning refer to the ability to learn on a sequence of tasks drawn from a potentially non-stationary distribution, and in particular, seek to do so while accelerating learning new tasks and without forgetting old tasks.
4. **Multi-Task Learning (MTL)** aims to jointly learn several related tasks, and benefits from the effect regularization due to parameter sharing and of the diversity of the resulting shared representation.
5. **Hyperparameter Optimization (HO)** is within the remit of meta-learning, in that hyperparameters such as learning rate or regularization strength can be included in the definition of ‘how to learn’.

### *Applications*

1. Multi-class image recognition
2. Few-shot object detection
3. Landmark Prediction
4. Few-Shot Object Segmentation
5. Image and Video Generation
6. Generative Models and Density Estimation

### *Challenges*

1. Diverse and multi-modal task distributions
2. Meta-generalization Meta-learning poses a new generalization challenge across tasks analogous to the challenge of generalizing across instances in conventional machine learning.
3. Task families: Many existing meta-learning frameworks, especially for few-shot learning, require task families for meta-training.

4. **Computation Cost & Many-shot:** A naive implementation of bilevel optimization is expensive in both time (because each outer step requires several inner steps) and memory (because reverse-mode differentiation requires storing the intermediate inner states)

## **Conclusions:**

I have presented a very high perspective information on architectures, optimization methods, and applications of multi-task and meta-learning.

I believe that the development of multi-task learning (and the related fields of meta-learning, transfer learning, and continuous/lifelong learning) is an important step towards developing artificial intelligence with more human-like qualities. In order to build machines that can learn as quickly and robustly as humans, we must create techniques for learning general underlying concepts that are applicable between tasks and applying these concepts to new and unfamiliar situations.

The field of meta-learning has seen a rapid growth in interest. This has come with some level of confusion, with regards to how it relates to neighboring fields, what it can be applied to, and how it can be benchmarked. In this survey, I have sought to the topics on a very high level to give the gist where the meta-learning stood till now and can go beyond.

I thank everyone who has visited the blog and request you to also go through the below reference for a detailed understanding of the topics

## **References:**

1. <https://arxiv.org/pdf/2004.05439.pdf>
2. <https://arxiv.org/pdf/2009.09796.pdf>