Meta-Learner with Linear Nulling

- An embedding network is combined with a linear transformer.
- The linear transformer carries out null-space projection on an alternative classification space.
- The projection space $\mathcal{M}$ is constructed to match the network output with a special set of reference vectors.

$\Phi = \{\phi_0, \phi_1, \phi_2, \phi_3\}$

$\mathcal{M} \leftarrow \text{null}(\{\nu_0, \nu_1, \nu_2, \nu_3\})$

$\nu_k = \phi_k' - g_k$: error vector

$\phi_k' = (N_c - 1)\phi_k - \sum_{j \neq k} \phi_j$

$\nu_k = g_k - \mathcal{M}$: error vector

$\phi_k' = \{\phi_0', \phi_1', \phi_2', \phi_3'\}$

Reference vectors

$\mathcal{M}$: Softmax

Distance measures

Classification

$\mathcal{M} = \text{null}(\{\nu_0, \nu_1, \nu_2, \nu_3\})$

$\mathcal{M}$ is constructed to match the network output with a special set of reference vectors.
Goal: Select models for a new dataset within time budget.

Given: Model performance and runtime on previous datasets.

Approach:
- **low rank** dataset-by-model collaborative filtering matrix
- **predict model runtime** using polynomials
- **classical experiment design** for cold-start
- missing entry imputation for model performance prediction

Performance:
- cold-start: high accuracy
- model selection: fast and perform well
Backpropamine: meta-learning with neuromodulated Hebbian plasticity

- Differentiable plasticity: meta-learning with Hebbian plastic connections
  - Meta-train both the baseline weight and plasticity of each connection to support efficient learning in any episode

- In nature, plasticity is under real-time control through neuromodulators
  - The brain can decide when and where to be plastic

- Backpropamine = Differentiable plasticity + neuromodulation
  - Make the rate of plasticity a real-time output of the network
  - During each episode, the network effectively learns by self-modification

- Results:
  - Solves tasks that non-modulated networks cannot
  - Improves LSTM performance on PTB language modeling task
Hyperparameter Learning via Distributional Transfer

Ho Chung Leon Law\(^1\), Peilin Zhao\(^2\), Junzhou Huang\(^2\) and Dino Sejdinovic\(^1\)

\(^1\)University of Oxford and \(^2\)Tencent AI Lab

**Goal (hyperparameter selection):**
Optimise \(f_{\text{target}}\) (target objective) w.r.t \(\theta\):

\[ \theta^*_{\text{target}} = \arg\max_{\theta \in \Theta} f_{\text{target}}(\theta) \]

**Scenario:**

- We have \(n\) potentially related tasks \(f^i, i = 1, \ldots n\)
- For these tasks, we have \(\{\theta^i_k, f^i(\theta^i_k)\}_{k=1}^{N_i}\) from past runs

**Method:**

- Assume training data \(D_i\) comes from distribution \(\mathcal{P}^i_{XY}\)
- Transfer information using embeddings of \(\mathcal{P}^i_{XY}\)
- Jointly model \(\theta, \mathcal{P}_{XY}\) and sample size \(s\)
Toward Multimodal Model-Agnostic Meta-Learning

Risto Vuorio¹, Shao-Hua Sun², Hexiang Hu² & Joseph J. Lim²

University of Michigan¹
University of Southern California²

The limitation of the MAML family

• One initialization can be suboptimal for multimodal task distributions.

Multi-Modal MAML

1. Model-based meta-learner computes task embeddings
2. Task embeddings are used to modulate gradient-based meta-learner
3. Gradient-based meta-learner adapts via gradient steps
1. Finds architecture for CNNs in ~0.25 days
2. Based on the idea of utility of individual nodes.
Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples

Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, Hugo Larochelle

- New benchmark for **few-shot classification**
- Two-fold approach:
  1. **Change the data**
     - Large-scale
     - Diverse
  2. **Change the task creation**
     - Introduce imbalance
     - Utilize class hierarchy for ImageNet

- Preliminary results on: baselines, Prototypical Networks, Matching Networks, and MAML.
- Leveraging data of multiple sources remains an open and interesting research direction!
Cell Search: the predefined skeleton ensures the simplest cell search can achieve 4.6% error with 0.4M params on CIFAR 10.

Key take-away: macro search can be competitive against cell search, even with simple random growing strategies, if the initial model is the same as cell search.

Cell Search: applies the found template on predefined skeleton.

Macro Search: learns all connections and layer types.
AutoDL challenge design and beta tests

Zhengying Liu*, Olivier Bousquet, André Elisseeff, Sergio Escalera, Isabelle Guyon,
Julio Jacques Jr., Albert Clapés, Adrien Pavao, Michèle Sebag, Danny Silver,
Lisheng Sun-Hosoya, Sébastien Tréguer, Wei-Wei Tu, Yiqi Hu, Jingsong Wang, Quanming Yao

Help Automating Deep Learning

Join the AutoDL challenge!
https://autodl.chalearn.org
Modular meta-learning in abstract graph networks for combinatorial generalization

Ferran Alet, Maria Bauza, A. Rodriguez, T. Lozano-Perez, L. Kaelbling

code&pdf:alet-etal.com

Combinatorial generalization: generalizing by reusing neural modules

Graph Neural Networks
Nodes tied to entities
- Objects
- Particles
- Joints

We introduce: Abstract Graph Networks
nodes are not tied to concrete entities

Modular meta-learning

Graph Element Networks
OmniPush dataset
Cross-Modulation Networks For Few-Shot Learning

Hugo Prol†, Vincent Dumoulin‡, and Luis Herranz†

†Computer Vision Center, Univ. Autònoma de Barcelona  
‡Google Brain

Key idea: allow support and query examples to interact at each level of abstraction.

Extending the feature extraction pipeline of Matching Networks:

☆ Channel-wise affine transformations: \( \text{FiLM}(x) = (1 + \gamma) \odot x + \beta \)
☆ Subnetwork G predicts the affine parameters \( \gamma \) and \( \beta \)

```
<table>
<thead>
<tr>
<th>Support set</th>
<th>Conv</th>
<th>BN</th>
<th>FiLM</th>
<th>ReLU</th>
<th>Max Pool</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Query set</td>
<td>Conv</td>
<td>BN</td>
<td>FiLM</td>
<td>ReLU</td>
<td>Max Pool</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
Large Margin Meta-Learning for Few-Shot Classification

The University of Hong Kong¹, The Hong Kong Polytechnic University²

Yong Wang¹, Xiao-Ming Wu², Qimai Li², Jiatao Gu¹, Wangmeng Xiang², Lei Zhang², Victor O.K. Li¹

Large Margin Principle

\[
\mathcal{L} = \mathcal{L}_{\text{softmax}} + \lambda \ast \mathcal{L}_{\text{large-margin}}
\]

Fig. 1: Large margin meta-learning. (a) Classifier trained without the large margin constraint. (b) Classifier trained with the large margin constraint. (c) Gradient of the triplet loss.

One Implementation: Triplet Loss

\[
\mathcal{L}_{\text{large-margin}} = \frac{1}{N_t} \sum_{i=1}^{N_t} \left[ \| f_\phi(x_i^s) - f_\phi(x_i^t) \|^2 - \| f_\phi(x_i^s) - f_\phi(x_j^d) \|^2 + m \right]_+.
\]

Case study
- Graph Neural Network (GNN)
- Prototypical Network (PN)

Analysis

After rearrangement:

\[
\mathcal{L}_{\text{large-margin}} = \frac{1}{N_t} \left( \sum_{x_i \in S_s} \| f_\phi(x_i) - f_\phi(x_s) \|^2 - \sum_{x_j \in S_d} \| f_\phi(x_i) - f_\phi(x_d) \|^2 \right) + \text{const.}
\]

The gradient:

\[
\frac{\partial \mathcal{L}_{\text{large-margin}}}{\partial f_\phi(x_i)} = \frac{2}{N_t} \left( \sum_{x_s \in S_s} (f_\phi(x_i) - f_\phi(x_s)) - \sum_{x_d \in S_d} (f_\phi(x_i) - f_\phi(x_d)) \right)
\]

\[
= - \frac{2|S_s|}{N_t} \left( \frac{1}{|S_s|} \sum_{x_s \in S_s} f_\phi(x_s) - f_\phi(x_i) \right) - \frac{2|S_d|}{N_t} \left( f_\phi(x_i) - \frac{1}{|S_d|} \sum_{x_d \in S_d} f_\phi(x_d) \right)
\]

\[
= - \frac{2|S_s|}{N_t} (c_s - f_\phi(x_i)) - \frac{2|S_d|}{N_t} (f_\phi(x_i) - c_d) .
\]

Features
- We implement and compare several of other large margin methods for few-shot learning.
- Our framework is simple, efficient, and can be applied to improve existing and new meta-learning methods with very little overhead.
Amortized Bayesian Meta-Learning
Sachin Ravi & Alex Beatson
Department of Computer Science, Princeton University

- Lot of progress in few-shot learning but under controlled settings
- In real world, relationship between training and testing tasks can be tenuous
  - Task-specific predictive uncertainty is crucial
- We present gradient-based meta-learning method for computing task-specific approximate posterior
- Show that method displays good predictive uncertainty on contextual-bandit and few-shot learning tasks

![Graph showing ECE and MCE]
The effects of negative adaptation in Model-Agnostic Meta-Learning

Tristan Deleu, Yoshua Bengio

- The advantage of meta-learning is well-founded under the assumption that the adaptation phase does improve the performance of the model on the task of interest.

- Optimization: maximize the performance after adaptation, performance improvement is not explicitly enforced.

$$\min_{\theta} \mathbb{E}_{T \sim p(T)} [\mathcal{L}(\theta'_{\mathcal{T}}; D'_{\mathcal{T}})]$$

- We show empirically that performance can decrease after adaptation in MAML. We call this negative adaptation.

- How to fix this issue? Ideas from Safe Reinforcement Learning.
Mitigating Architectural Mismatch During the Evolutionary Synthesis of Deep Neural Networks

Audrey G. Chung, Paul Fieguth, Alexander Wong

- *Evolutionary deep intelligence* for increasingly efficient networks
- Preliminary study into the effects of architectural alignment
- Like-with-like mating policy via gene tagging system
- Resulting networks are comparable:
  - Restricts search space exploration?
  - Compensated with training epochs?
  - ???
Evolvability ES: Scalable Evolutionary Meta-Learning

By Alexander Gajewski, Jeff Clune, Kenneth O. Stanley, and Joel Lehman

- Evolvability ES is a meta-learning algorithm inspired by Evolution Strategies [1]
- Surprisingly, Evolvability ES finds parameters such that at test time, random perturbations result in diverse behaviors
- In a simulated Ant locomotion domain, adding Gaussian noise to the parameters results in policies which move in many different directions

Consolidating the Meta-Learning Zoo
A Unifying Perspective as Posterior Predictive Inference


► Meta-learning: Learns how to learn a classifier or regressor for each new task.

► Unifies: MAML, Meta-LSTM, Prototypical networks, and Conditional Neural Processes are special cases.


► Efficient: Test-time requires only forward passes, no gradient steps are needed.

► Versatile: Robust classification accuracy as shot and way are varied at test-time.

► High quality 1-shot view reconstruction:
Deep Online Learning via Meta-Learning: Continual Adaptation for Model-Based RL
Anusha Nagabandi, Chelsea Finn, Sergey Levine

Can we use meta-learning for effective online learning?

Our method can:
- Reason about non-stationary latent distributions over tasks.
- Recall past tasks