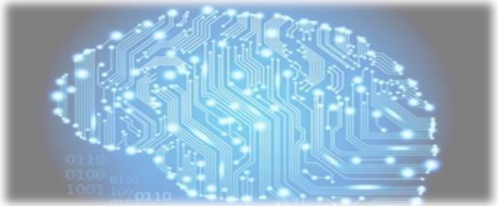
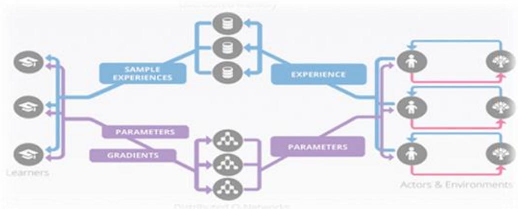



M-NAS: Meta Neural Architecture Search

AutoML 연구실
석사과정 문아성

21.05.11

M-NAS: Meta Neural Architecture Search

- M-NAS: Meta Neural Architecture Search

Wang, Jiaxing, et al. "M-nas: Meta neural architecture search."

Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. No. 04. 2020.

- Universal NAS method

  - Task-aware search

  - Transferable knowledge modulation

M-NAS Framework

- Task-aware Architecture Generation
- Transferable Knowledge Modulation
- Fast weights estimation

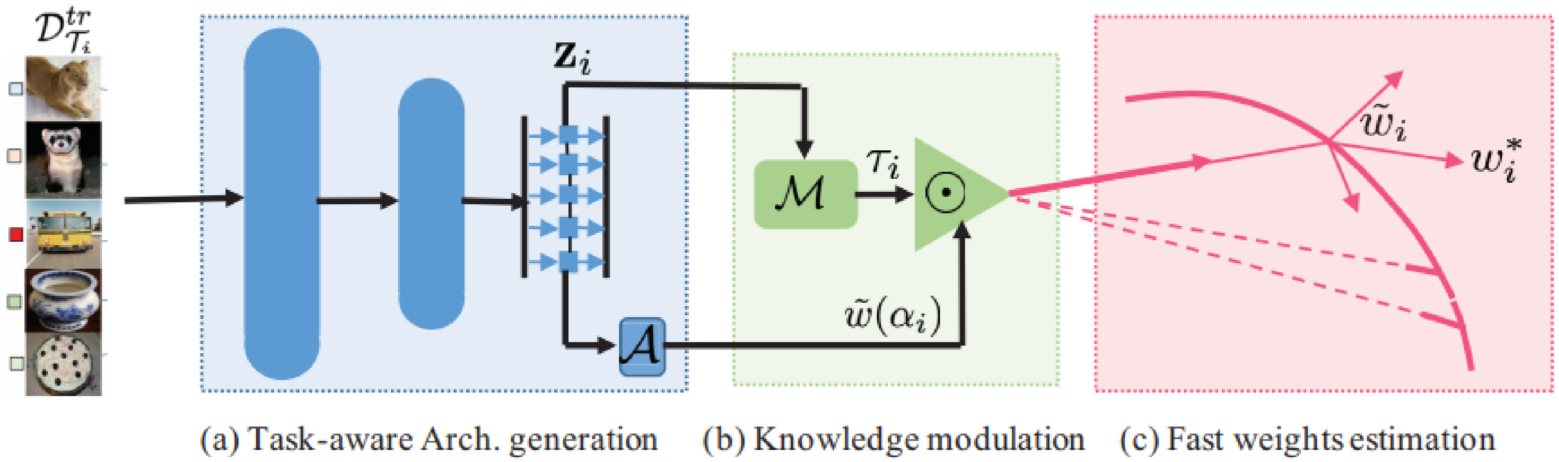
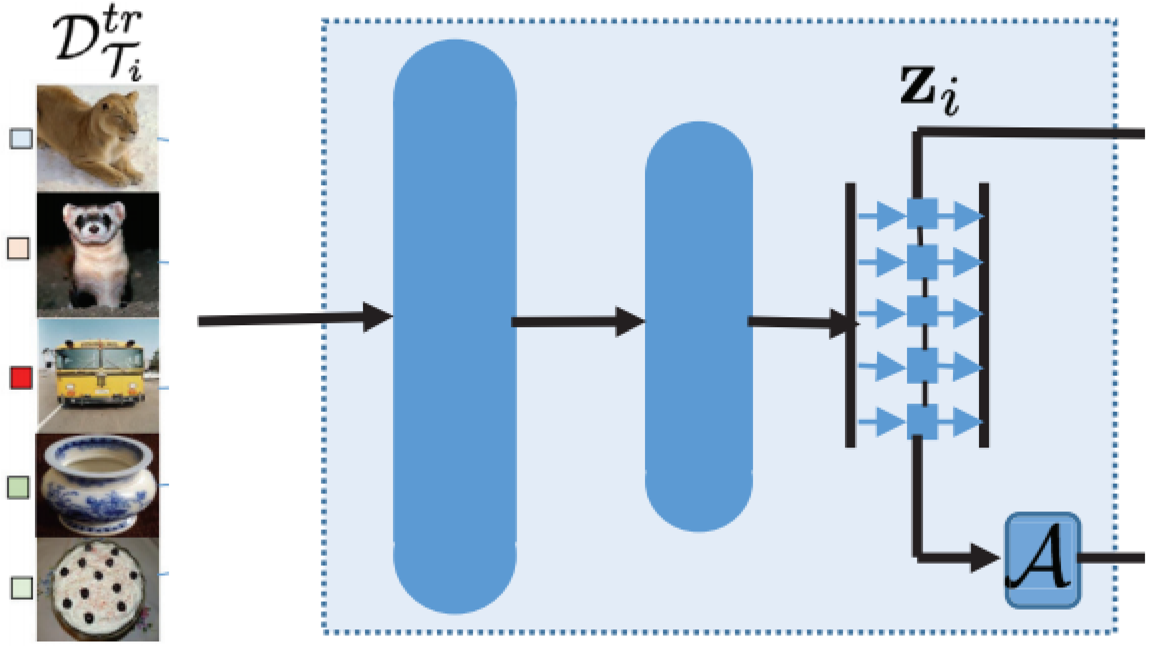


Figure 1: The framework of the proposed M-NAS involving three essential stages. (a) Task conditional architecture generating: we extract representation for the task $\mathcal{T}_i$ using an recurrent auto-encoder, which is then used to generate a task-aware model architecture α_i. (b) Transferable knowledge modulation: we tailor the globally shared meta-parameters associated with α_i, denoted as $\tilde{w}(\alpha_i)$, to a specific task $\tilde{w}_i$. (c) Fast optimal weights estimation: Given an architecture α_i and the modulated meta-parameters $\tilde{w}_i$, apply a few gradient descent steps to get an estimation of the optimal weights $w_i^*(\alpha_i)$.

Task-aware Architecture Generation

- $z_{i,j} = RNN_{enc}(\mathcal{F}(x_{i,j}^{tr},\ y_{i,j}^{tr}),\ z_{i,j-1})$
- $d_{i,j} = RNN_{dec}(z_{i,j},\ d_{i,\ j+1})$
- $\mathcal{L}_r(\mathcal{D}_{\mathcal{T}_i}^{tr}) = \sum_{j=1}^{n^{tr}} ||d_{i,j} - \mathcal{F}(x_{i,j}^{tr},\ y_{i,j}^{tr})||_2^2$
- $z_i = \frac{1}{n^{tr}} \sum_j^{n^{tr}} (z_{i,j})$
- $\alpha_i = \mathcal{A}(z_{i,j})$



- $\{x_{i,j}^{tr},\ y_{i,j}^{tr}\}_{j=1}^{n^{tr}} \in \mathcal{D}_{\mathcal{T}_i}^{tr}$
- i: *task* j: *image*
- $RNN_{enc},\ RNN_{dec}$: *auto-encoder*
- z_i: *Task embedding vector*
- $\mathcal{A}$: *LSTM controller or trainable architecture parameters*

Transferable Knowledge Modulation

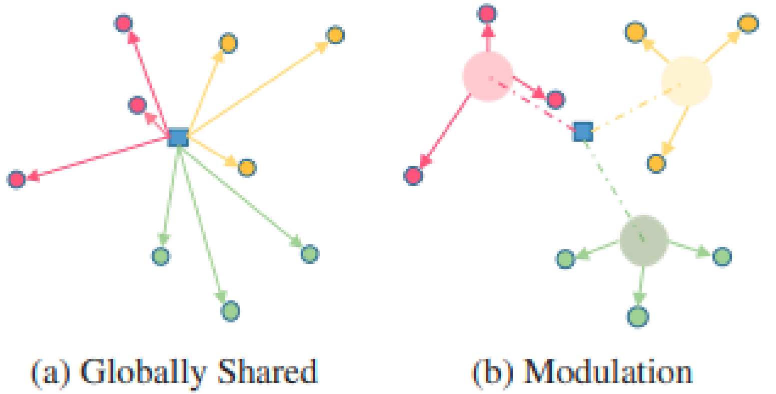


Figure 2: Task specific modulation on the meta-parameters. Circles represent optimal solutions for tasks, which forms a distribution with three modalities. (a) Globally Shared meta-parameters: model is having hard time to adapt to solutions when tasks are heterogeneous. (b) Meta-parameters with task-specific modulation. Meta-weights are first modulated to be closer to the task clusters.

- $\mathcal{T}_i = \mathcal{M}(z_i;\ \mathcal{W}_\mathcal{M})$
- $\widetilde{w_i} = \mathcal{T}_i \odot \tilde{w}$

- $\tilde{w}$: *a set of meta-parameters*
- $\mathcal{T}_i$: *task cluster*
- $\mathcal{M} \sim \mathcal{W}_\mathcal{M}$: *Modulator, weights*

M-NAS: Meta Neural Architecture Search

Fast weights estimation

- $w_i^* = \tilde{w}_i - \rho_{\text{inner}} \nabla_{\tilde{w}_i} \mathcal{L}(\tilde{w}_i, \mathcal{D}_{\mathcal{T}_i}^{tr})$
- $\min\limits_{\Theta} \sum\limits_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}\left(w_i^*(\alpha_i), \mathcal{D}_{\mathcal{T}_i}^{val}\right) + \beta \mathcal{L}_r(\mathcal{D}_{\mathcal{T}_i}^{tr})$

Algorithm 1 M-NAS: Meta Neural Architecture Search

Require: Meta-train dataset $\mathcal{D}_{\text{meta-train}}$, learning rates ρ_{inner}, ρ_{outer}, hyper-parameter β

Ensure:

Task representation extractor: $\mathcal{E}$

Task-aware architecture controller: $\mathcal{A}$

Shared meta-parameters on the big one-shot model $\tilde{w}$

Task-specific modulator $\mathcal{M}$

```
Randomly initialize Θ = {W_E, W_A, W_m, w̃}
while not done do
    Sample batch of tasks {T} in D_meta-train.
    for T_i ∈ T do
        Get the task representation z_i and generate an candidate architecture α_i.
        Modulate the meta-parameters with w̃_i = τ_i ⊙ w̃.
        Approximate the optimal weights with Equation (8).
    end for
    Update Θ by optimizing Equation (9).
end while
```

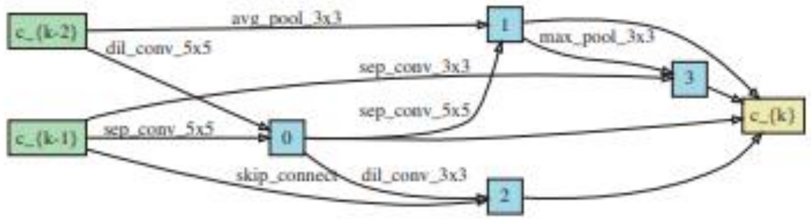
M-NAS: Meta Neural Architecture Search

Experimental Settings

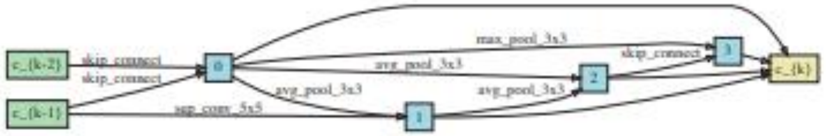
- Datasets
  - Caltech-UCSD Birds200-2011 (Bird)
  - Describable Textures Dataset (Texture)
  - Fine-Grained Visual Classification of Aircraft (Aircraft)
  - FGVCx-Fungi (Fungi)

- Search Space
  - 3 × 3 and 5 × 5 separable convolutions
  - 3 × 3 and 5 × 5 dilated separable convolutions
  - 3 × 3 max pooling, average pooling
  - identity and zero mapping

Experiment Result Architecture



(a) Normal cell



(b) Reduction cell

Figure 5: Architecture searched when learning a shared single base learner in 5-way 1-shot setting of Multi-datasets. The searched architecture contains 27.0K parameters

M-NAS: Meta Neural Architecture Search

Experiment Results

	Model	Params.	Bird	Texture	Aircraft	Fungi	Average
5 way 1 shot	MAML	32.9 K	53.94 ± 1.45%	31.66 ± 1.31%	51.37 ± 1.38%	42.12 ± 1.36%	44.77%
	MT-Net	32.9 K	58.72 ± 1.43%	32.80 ± 1.35%	47.72 ± 1.46%	43.11 ± 1.42%	45.59%
	BMAML	32.9 K	54.89 ± 1.48%	32.53 ± 1.33%	53.63 ± 1.37%	42.50 ± 1.33%	45.89%
	MUMOMAML	32.9 K	56.82 ± 1.49%	33.81 ± 1.36%	53.14 ± 1.39%	42.22 ± 1.40%	46.50%
	HSML	32.9 K	58.22 ± 1.48%	33.30 ± 1.36%	55.35 ± 1.38%	42.68 ± 1.40%	47.39%
	M-NAS-NM	31.6 K*	52.37 ± 1.50%	32.49 ± 1.48%	48.53 ± 1.42 %	40.68 ± 1.45%	43.52%
	M-NAS-Shared	27.0 K	57.13 ± 1.43%	**34.75 ± 1.33**%	56.32 ± 1.38%	43.33 ± 1.39%	**47.88**%
	M-NAS	29.5 K*	**58.76** ± **1.47**%	34.68 ± 1.36%	**57.13** ± **1.41**%	**43.71** ± **1.41**%	**48.57**%
5 way 5 shot	MAML	32.9 K	68.52 ± 0.79%	44.56 ± 0.68%	66.18 ± 0.71%	51.85 ± 0.85%	57.78%
	MT-Net	32.9 K	69.22 ± 0.75%	46.57 ± 0.70%	63.03 ± 0.69%	53.49 ± 0.83%	58.08%
	BMAML	32.9 K	69.01 ± 0.74%	46.06 ± 0.69%	65.74 ± 0.67%	52.43 ± 0.84%	58.31%
	MUMOMAML	32.9 K	70.49 ± 0.76%	45.89 ± 0.69%	67.31 ± 0.68%	53.96 ± 0.82%	59.41%
	HSML	32.9 K	71.31 ± 0.75%	47.11 ± 0.71%	**71.51** ± **0.69** %	54.30 ± 0.79%	61.06 %
	M-NAS-NM	31.6 K*	67.78 ± 0.81%	45.81 ± 0.72%	66.35 ± 0.73%	50.73 ± 0.86%	57.67 %
	M-NAS-Shared	27.0 K	**72.24** ± **0.75**%	47.15 ± 0.71%	70.87 ± 0.67%	55.23 ± 0.80%	**61.37** %
	M-NAS	29.5 K*	72.22 ± 0.78%	**48.17** ± **0.74**%	71.31 ± 0.70%	**55.85** ± **0.82**±	**61.89** %

Table 2: Comparison between M-NAS and other gradient-based meta-learning methods on the 5-way, 1-shot/5-shot image classification problem, averaged over 250 tasks for each dataset. Accuracy ± 95% confidence intervals are reported.