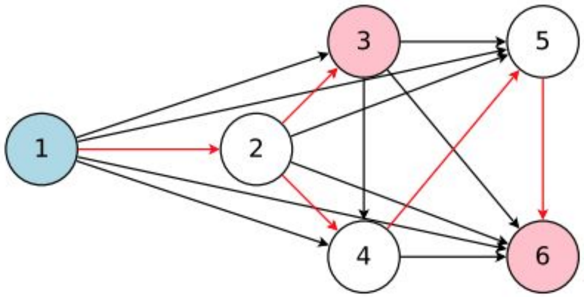
Overcoming Multi-Model Forgetting

Y. Benyahia, K. Yu, K. Bennani-Smires, M. Jaggi, A. Davison, M. Salzmann, C. Musat

The Weight Sharing

In One of the first NAS papers using Reinforcement Learning, Zoph et Al. (Google) used more than **800 gpus** in parallel for **two weeks**.

Weight Sharing was introduced in NAS to **speed up** the process



Efficient Neural Architecture Search (Pham et al.)

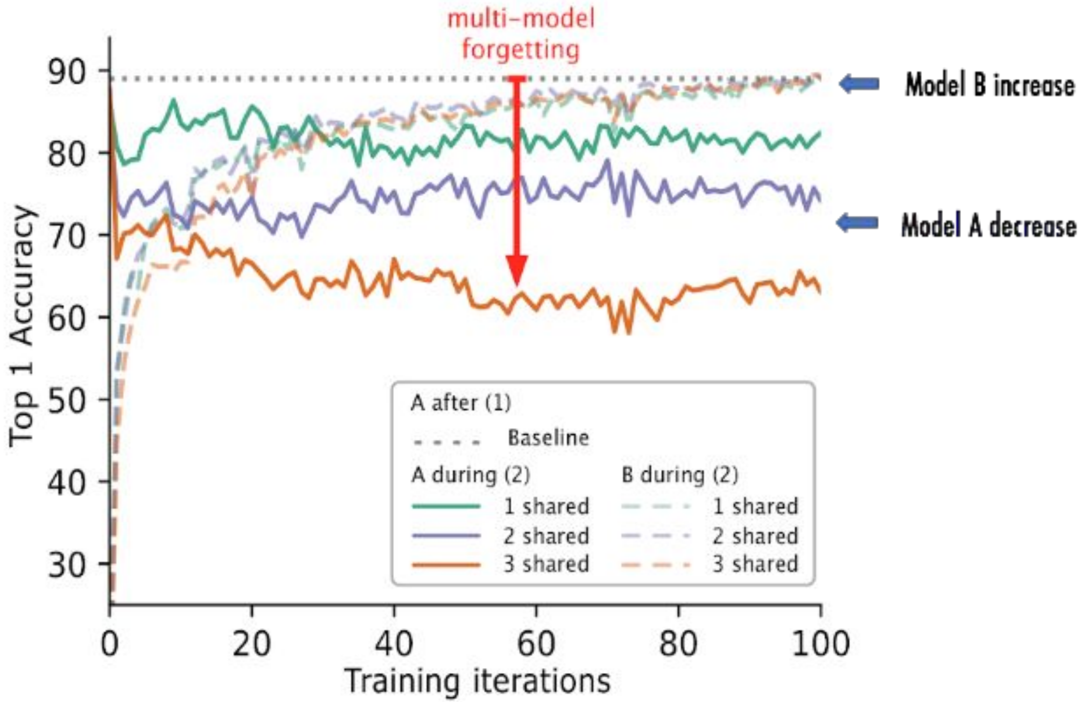
Assumptions

Our hypothesis:

1. Weight-sharing **can negatively affect architectures**.

2. If justified, this can lead to a **wrong evaluation** of candidates in **NAS**, making the evaluation phase **closer** to **random**

Multi-Model Forgetting



Study of Weight-Sharing

Simple scenario of two models **sharing** parameters:

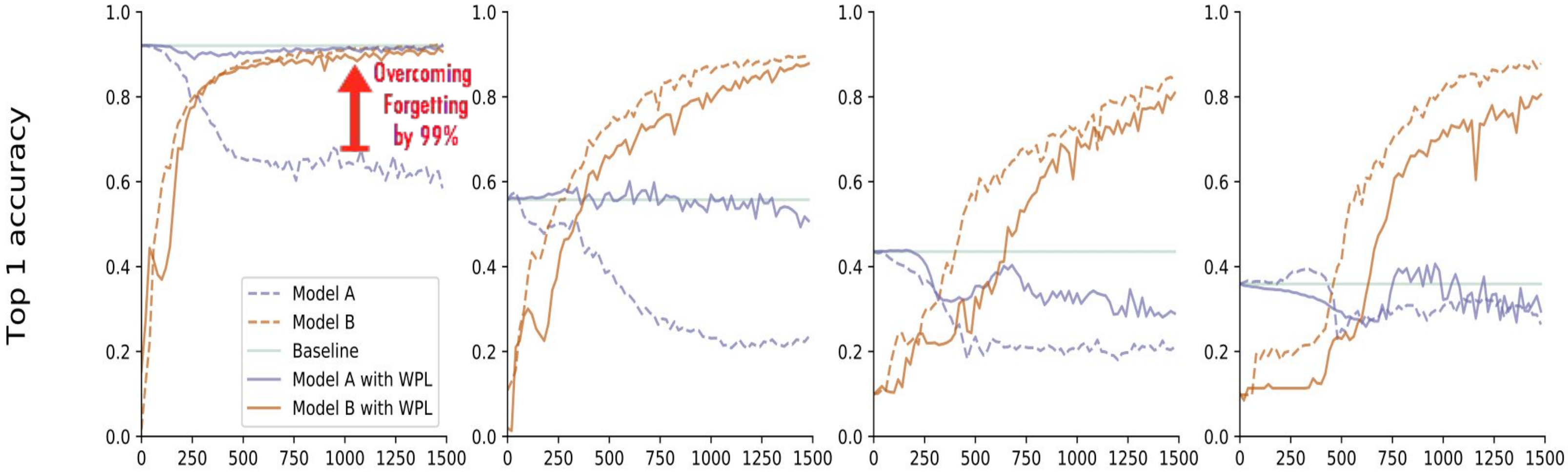
$f_1(\mathcal{D}; \theta_1, \theta_s)$ and $f_2(\mathcal{D}; \theta_2, \theta_s)$

Assume that we have access to the optimal parameters $(\hat{\theta}_1, \hat{\theta}_s)$ of the first model $f_1(\mathcal{D}; \theta_1, \theta_s)$

Maximizing the posterior distribution $p(\theta \mid \mathcal{D})$, $\theta = (\theta_1, \theta_2, \theta_s)$

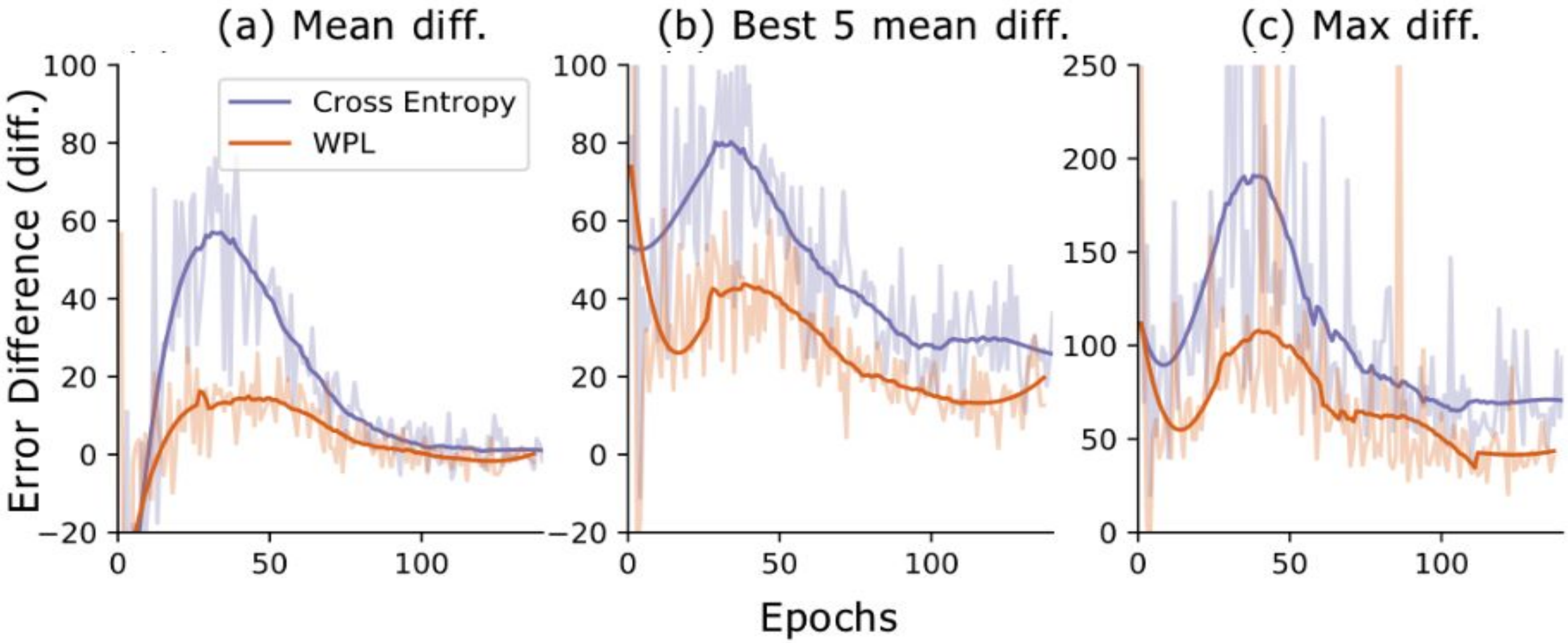
$$\mathcal{L}_{\mathrm{WPL}}(\theta_2, \theta_s) = \underbrace{\mathcal{L}_2(\theta_2, \theta_s)}_{\textbf{Cross-entropy loss}} + \underbrace{\frac{\lambda}{2}(\|\theta_s\|^2 + \|\theta_2\|^2)}_{\textbf{L2 regularization}} + \boxed{\underbrace{\frac{\alpha}{2} \sum_{\theta_{s_i} \in \theta_s} F_{\theta_{s_i}} (\theta_{s_i} - \hat{\theta}_{s_i})^2}_{\textbf{Weight importance}}}$$

Experiments on Two Models



- **WPL** reduces multi-model forgetting
- **WPL** have a **minimal** effect on the learning of the second model

ENAS on PTB

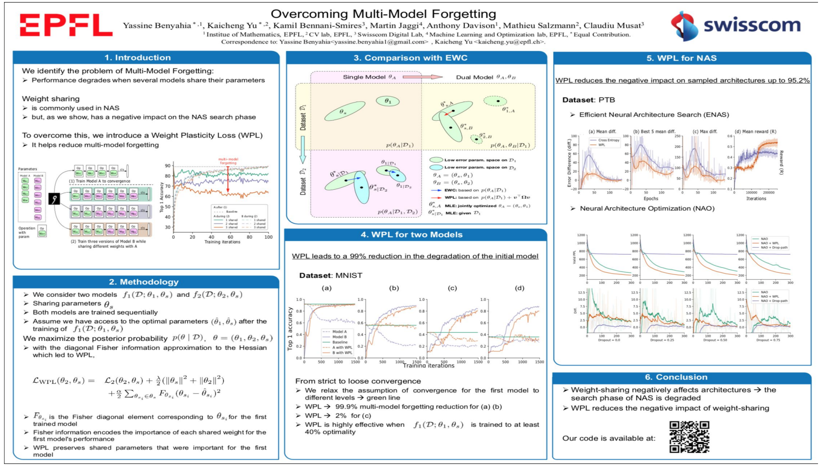


(a) Mean diff.
(b) Best 5 mean diff.
(c) Max diff.
Cross Entropy
WPL
Error Difference (diff.)
Epochs
100
80
60
40
20
0
−20
250
200
150
100
50
0
50
100

Summing up

To recap, our main contributions are:

1. **Weight Sharing** negatively impacts NAS
2. **Weight Sharing** can cause the **search phase** in NAS to become closer to random
3. **WPL** reduces Multi-Model Forgetting



Pacific Ballroom #19

(6:30pm - 9pm)