One Initialization to Rule them All: Fine-tuning via Explained Variance Adaptation

Fabian Paischer*[1], Lukas Hauzenberger*[1], Thomas Schmied[1], Benedikt Alkin[1,3] Marc Peter Deisenroth[2] Sepp Hochreiter [1,3]

[1] ELLIS Unit, LIT AI Lab, Institute for Machine Learning, JKU Linz, Austria

[2] University College London, [3] NXAI GmbH, Linz, Austria

MOTIVATION

Existing works either focus on initialization *or* adaptive ranks. EVA **combines** data-driven initialization with adaptive rank allocation to enhance performance and reduce parameters.

Method	Initialization	Adaptive ranks
LoRA	Random	✗
AdaLoRA	Random	✓
PiSSA	Weight-driven	✗
OLoRA	Weight-driven	✗
LoRA-GA	Data-driven	✗
EVA (Ours)	Data-driven	✓

TRY OUT EVA

PEFT

Code

Paper

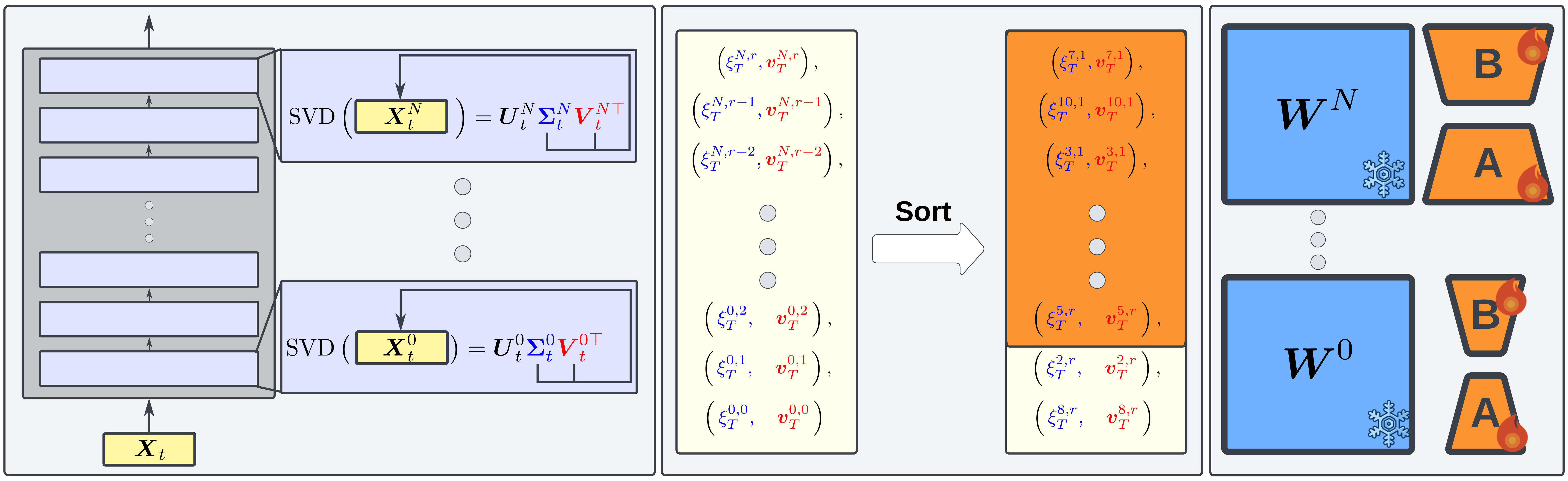
INITIALIZATION SPEED

Initialization	Method	Initialization	Training	% of Training
Weight-driven	PiSSA	7.43	482.67	1.5
	OLoRA	0.3	482.67	0.1
Data-driven	LoRA-GA	11.7	482.67	2.4
	$EVA_{bs=16}$	3.3	482.67	0.7
	$EVA_{bs=8}$	1.38	482.67	0.3
	$EVA_{bs=4}$	1.17	482.67	0.2

METHOD

Explained **V**ariance **A**daptation (EVA) initializes LoRA in a *data-driven* manner in two steps:

1. SVD on minibatches of activation vectors and initialize $\boldsymbol{A}$ matrices with right-singular vectors
2. re-distribute ranks to maximize amount of information about downstream data across weights



RESULTS