

Visual Analysis of the Impact of Neural Network Hyper-Parameters

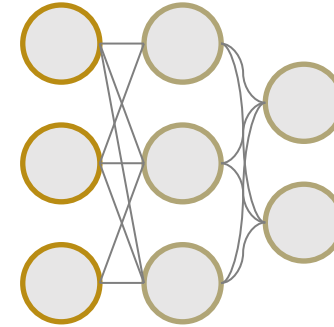
Daniel Jönsson¹, Gabriel Eilertsen¹, Hezi Shi², Jianmen Zheng², Anders Ynnerman¹, and Jonas Unger¹

¹Linköping University, Sweden

²Nanyang Technological University, Singapore

Scenario

- About to train a Convolutional Neural Network
- Which hyper-parameters should be chosen?



Hyper-parameters are all the training variables set manually with a pre-determined value before starting the training.

Finding the optimal parameter combination in this high dimensional space is a **non-trivial** problem



Compute intensive

Hyper-
parameter selection

Evaluate
performance

Full training



Related work

Manual

Primarily use **prior knowledge/best practices** to select good parameters

Human-in-the-loop

Interactive visual analysis, e.g.,
image analysis parameter analysis [Torsney-Weir et al., TVCG, 2012]
or
hyper-parameter tuning [Li et al., IEEEVIS Workshop MLUI, 2018]

Automatic

Optimization methods, e.g.,
random search [Bergstra et al., Journal of Machine Learning Research, 2012]
or
Bayesian optimization [Johan Mockus, Bayesian approach to global optimization: theory and applications, 2012]



Goal

Better **general** understanding
of the performance relationship
between CNN hyper-parameters



Starting point:

Classifying the Classifier: Dissecting the weight space of neural networks,



Random selection of hyper parameters for 13 000 combinations

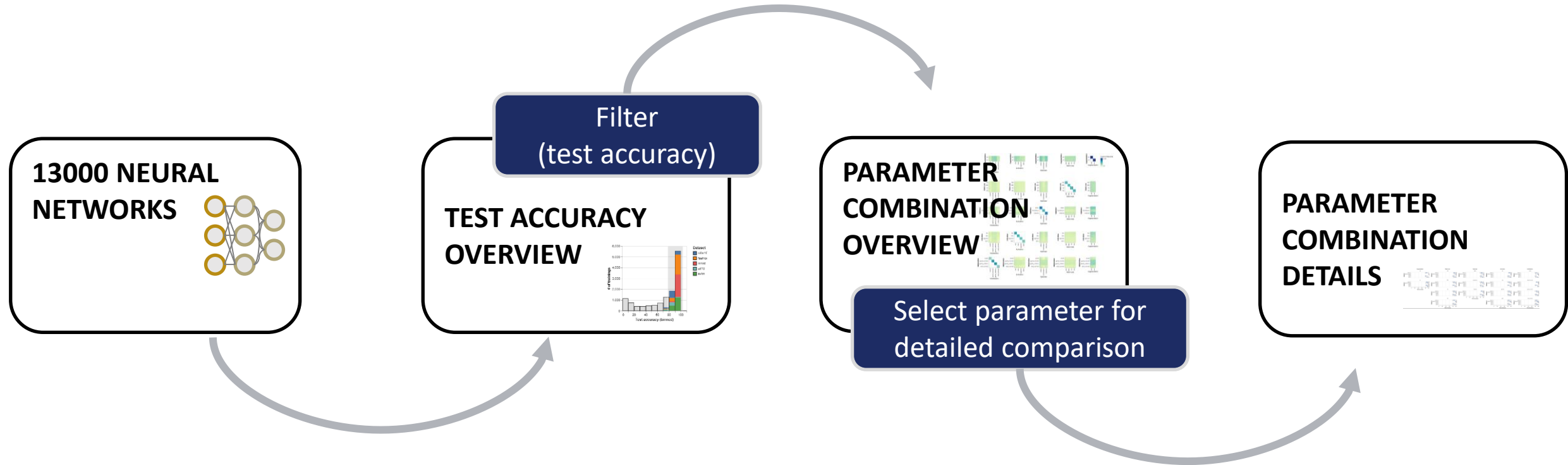


Hyper-parameters

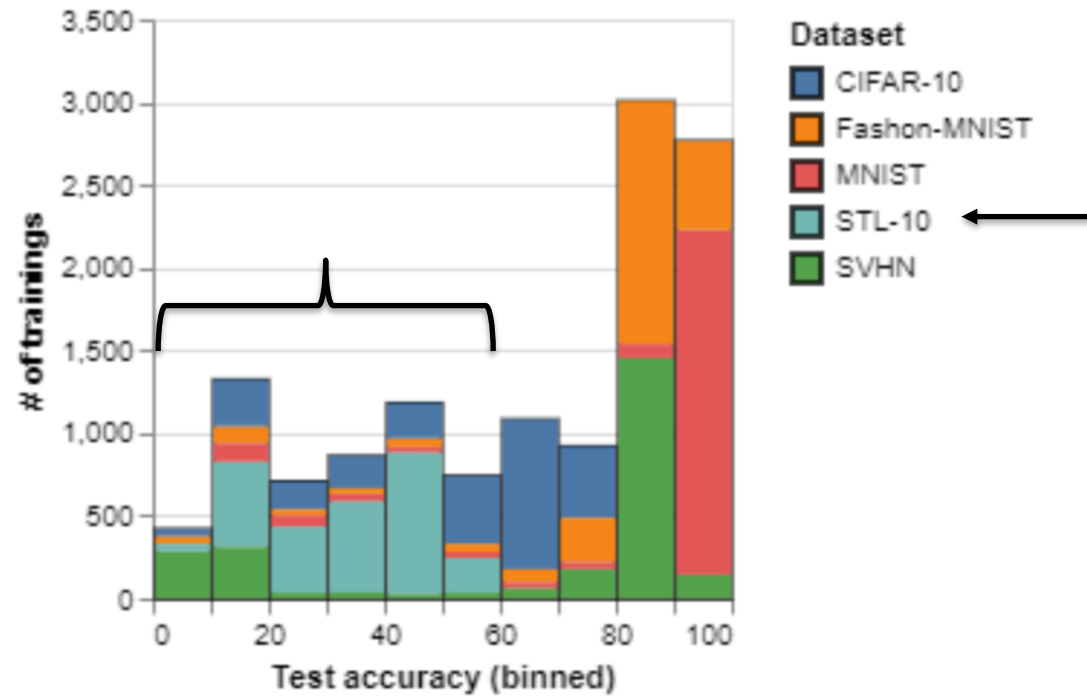
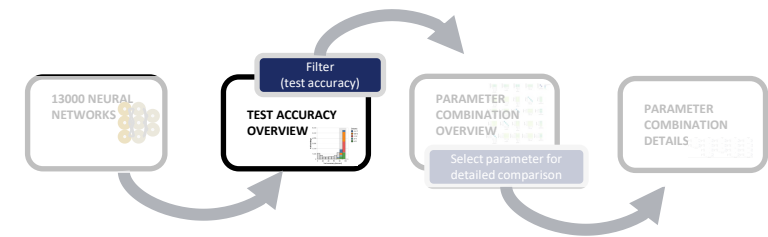
Dataset	MNIST [LBB*98], CIFAR-10 [KH09], SVHN [NWC*11], STL-10 [CNL11], Fashion-MNIST [XRV17]
Batch size	32, 64, 128, 256
Augmentation	Off, On
Optimizer	ADAM [KB14], RMSProp [HSS12], Mo- mentum SGD
Activation	ReLU [NH10], ELU [CUH15], Sigmoid, TanH
Initialization	Constant, Random normal, Glorot uniform, Glorot normal [GB10]



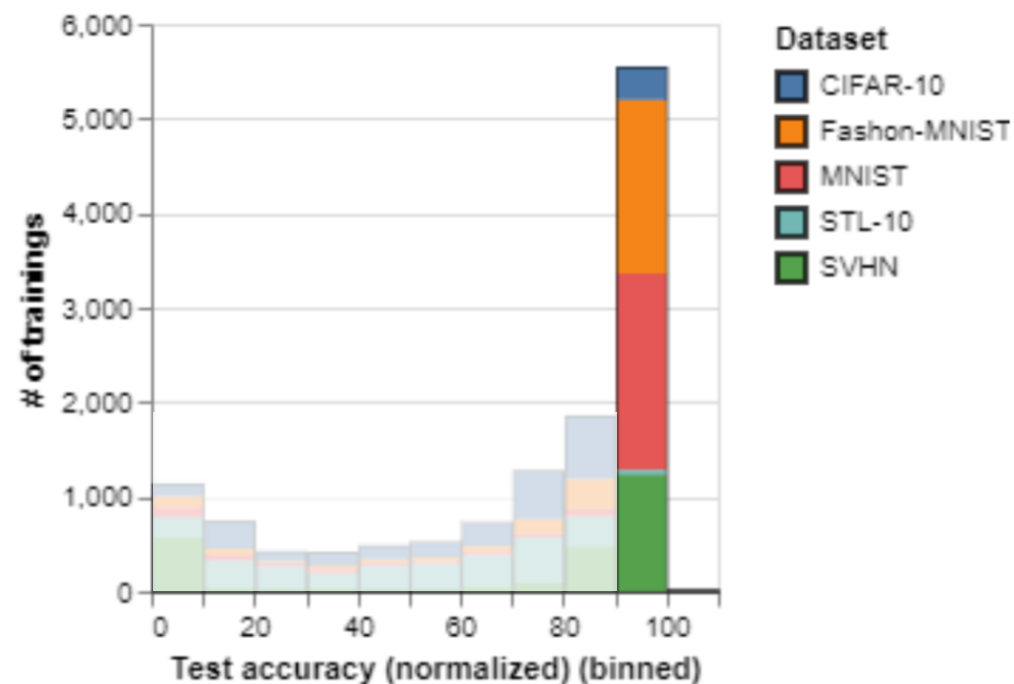
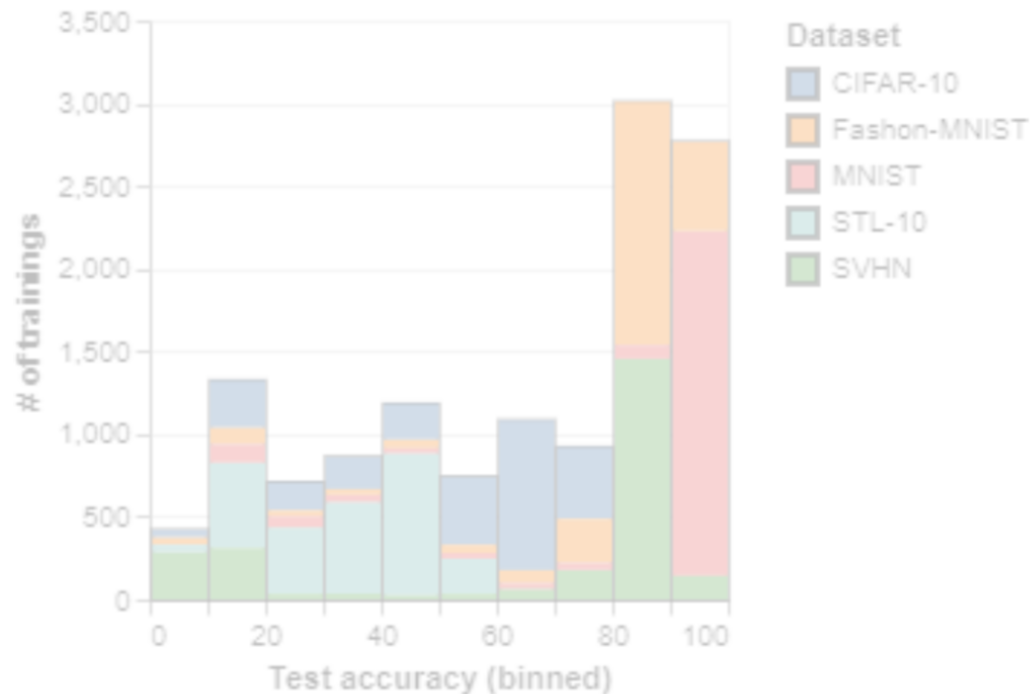
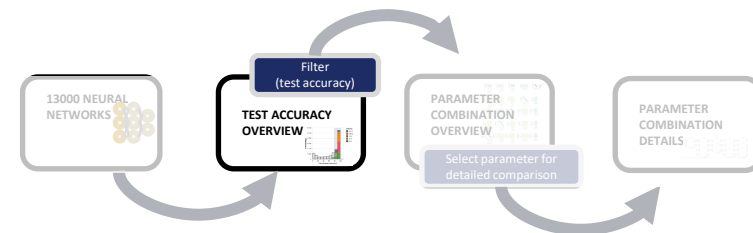
A visual analysis approach



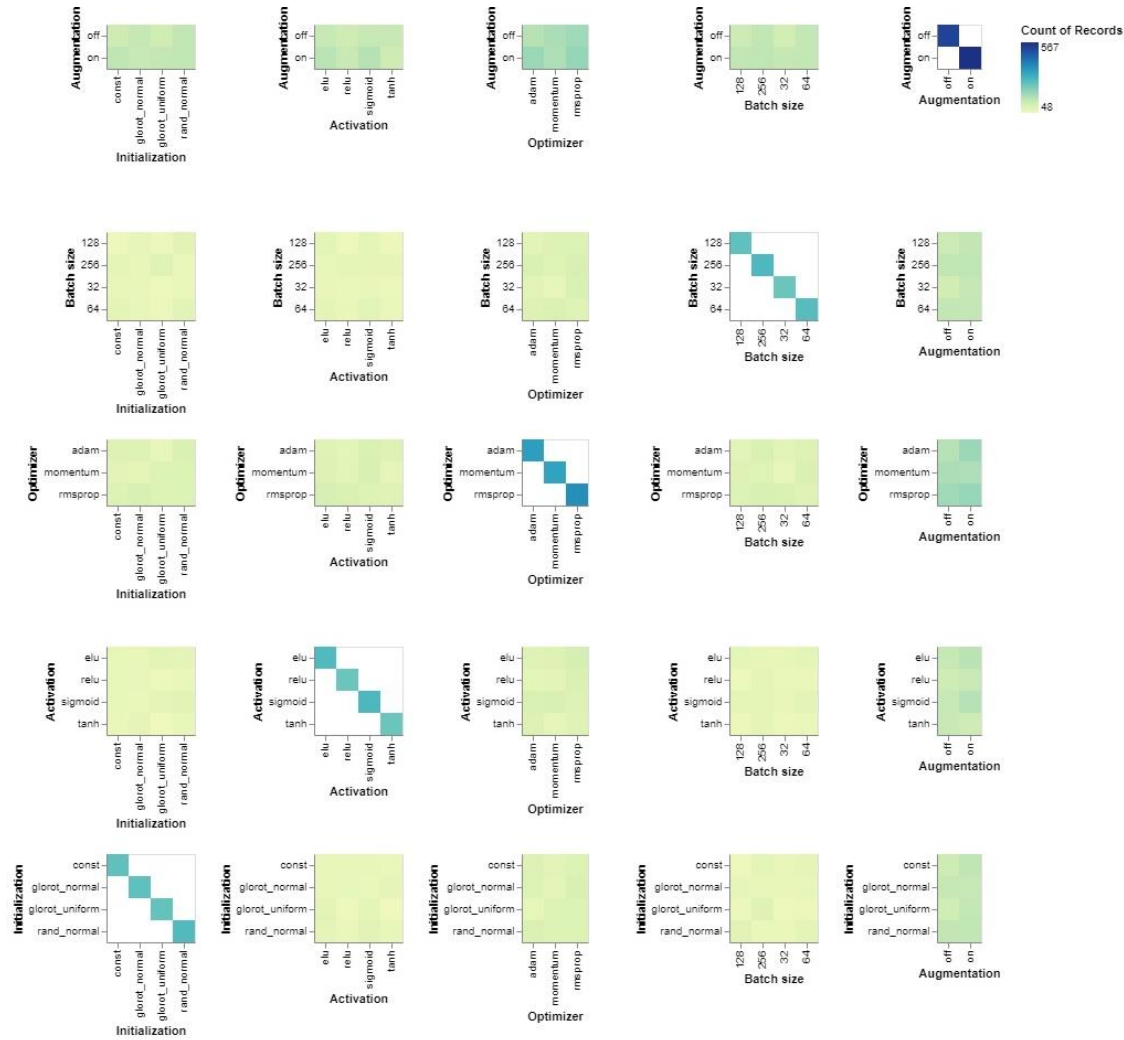
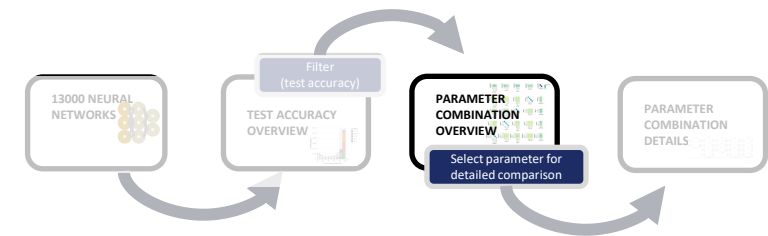
Primary properties: data set and performance



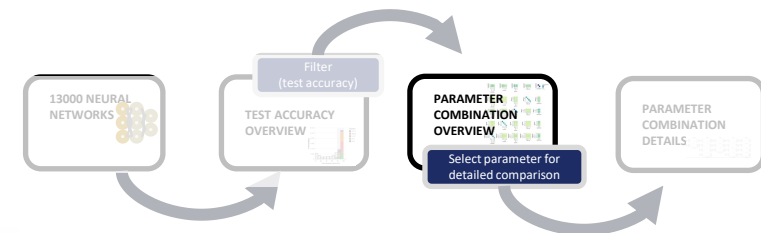
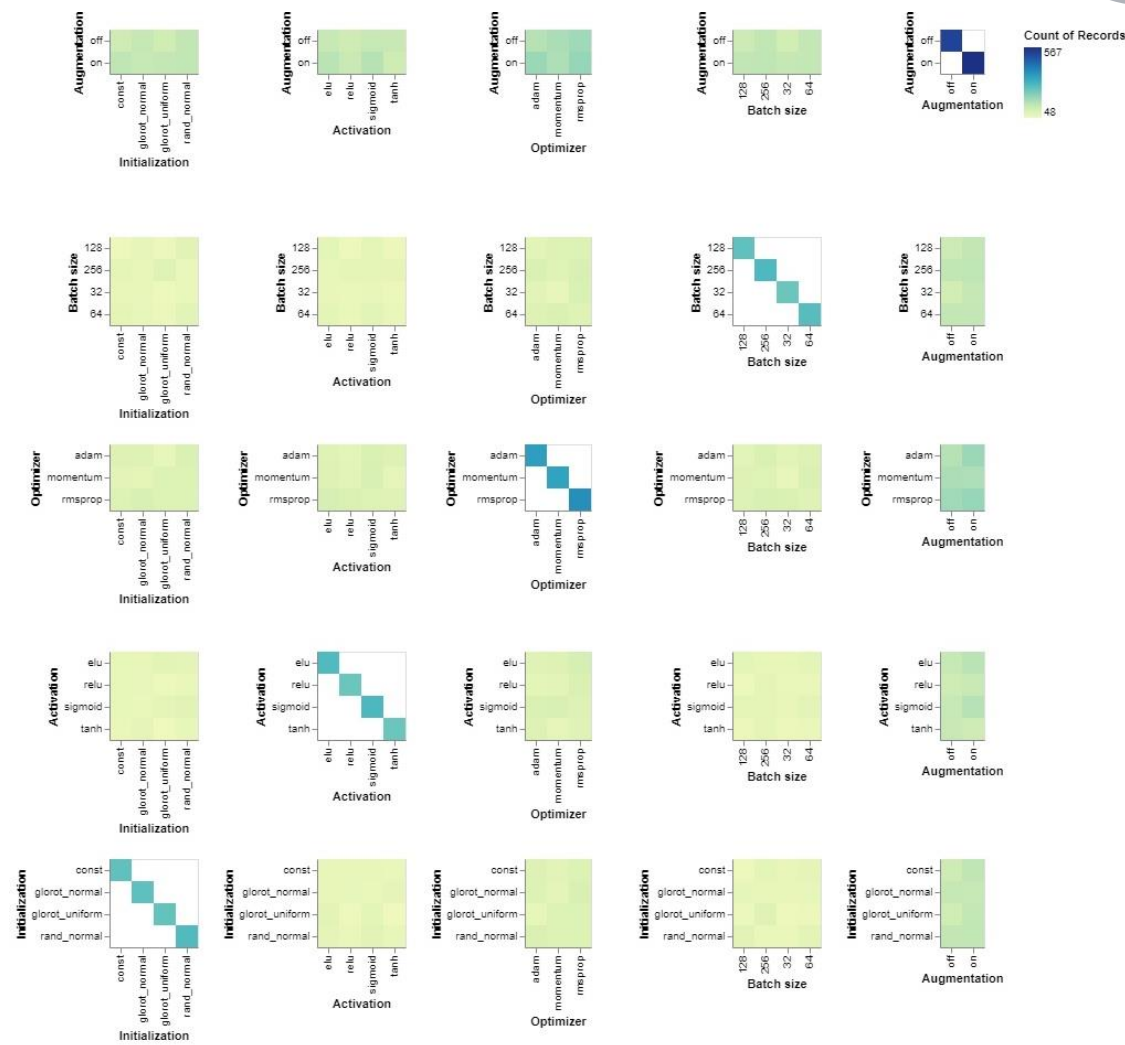
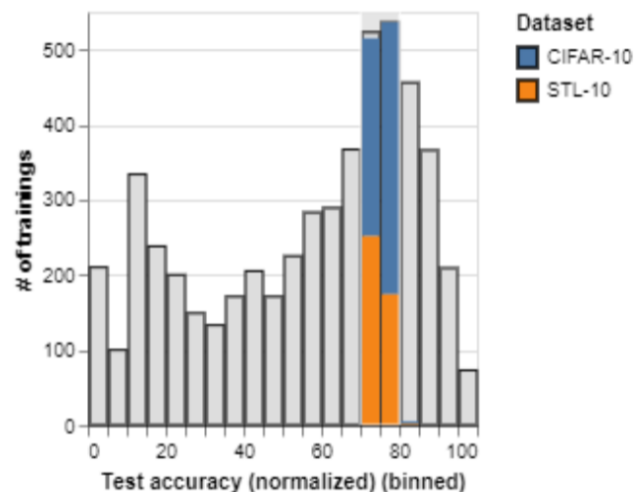
Problem: Varying data set complexity



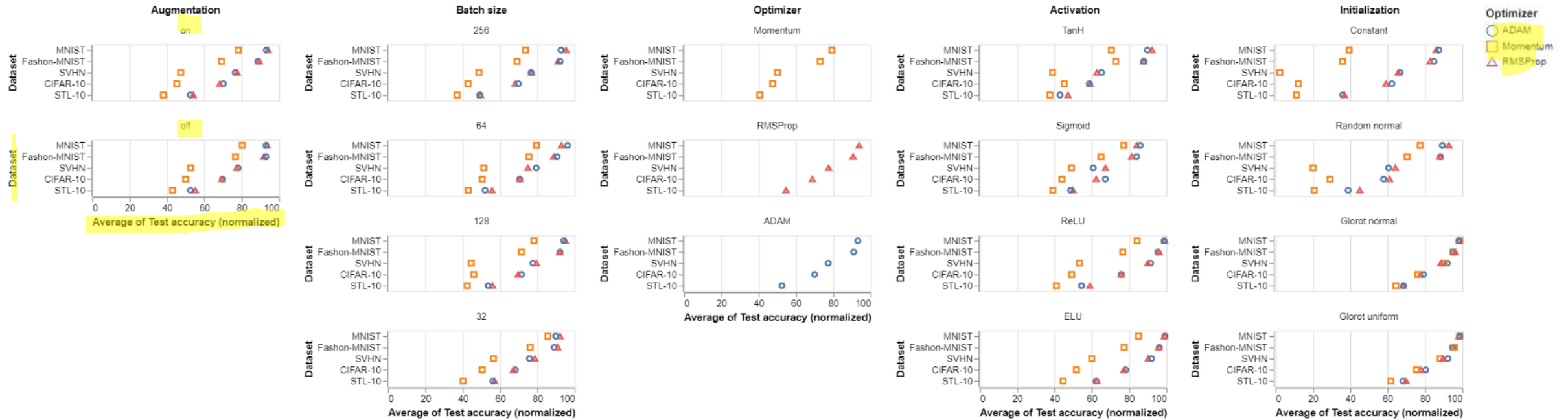
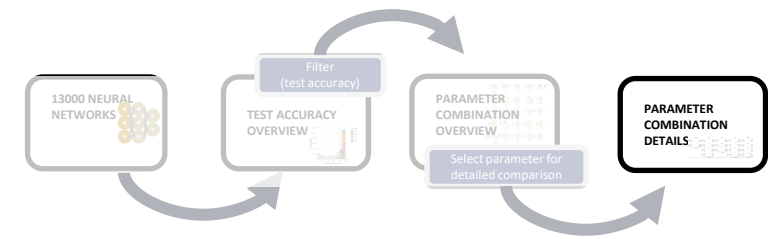
Parameter comparison overview



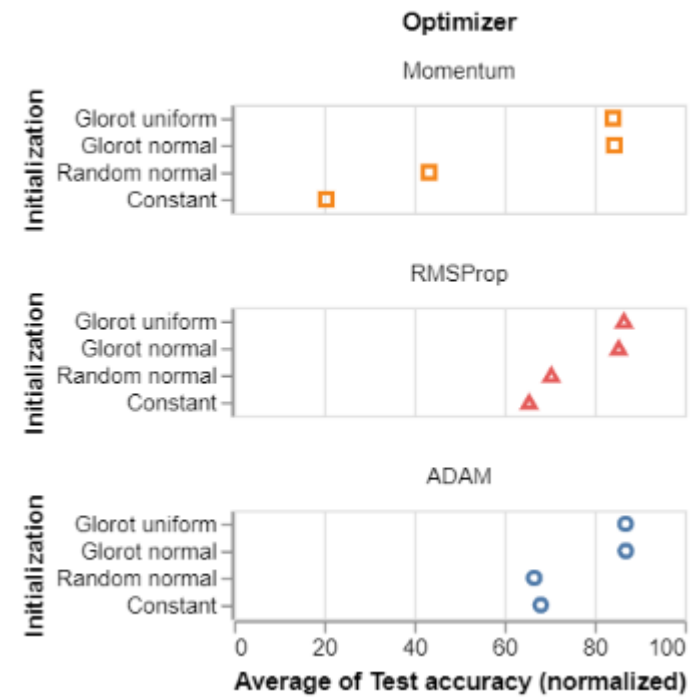
Combined with interaction



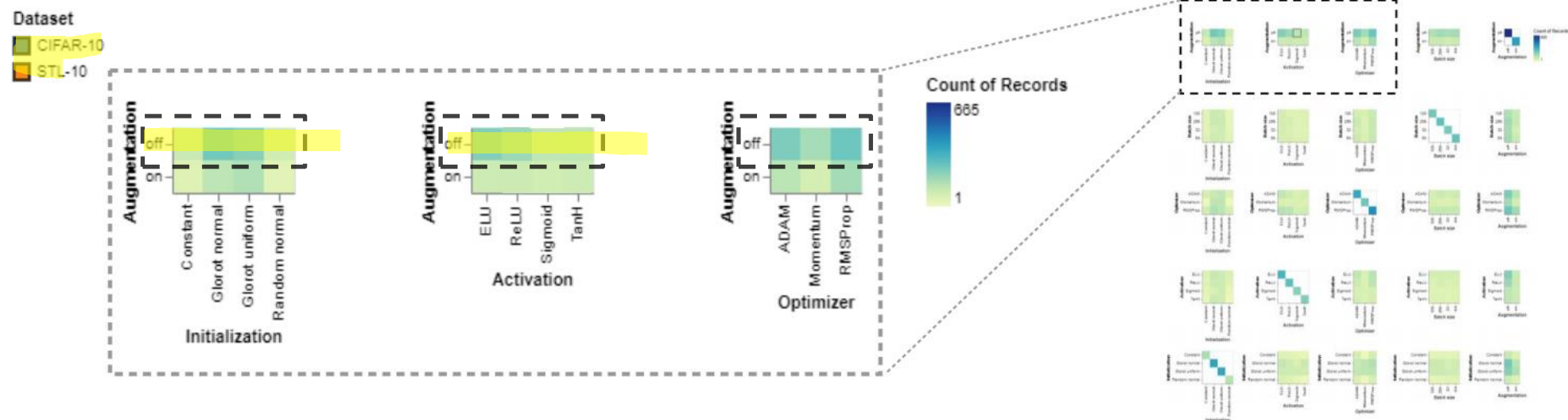
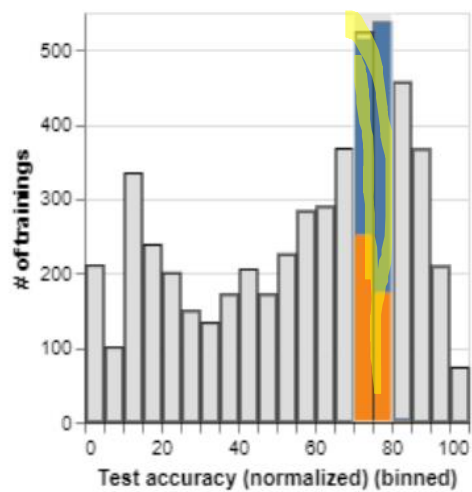
Detailed view



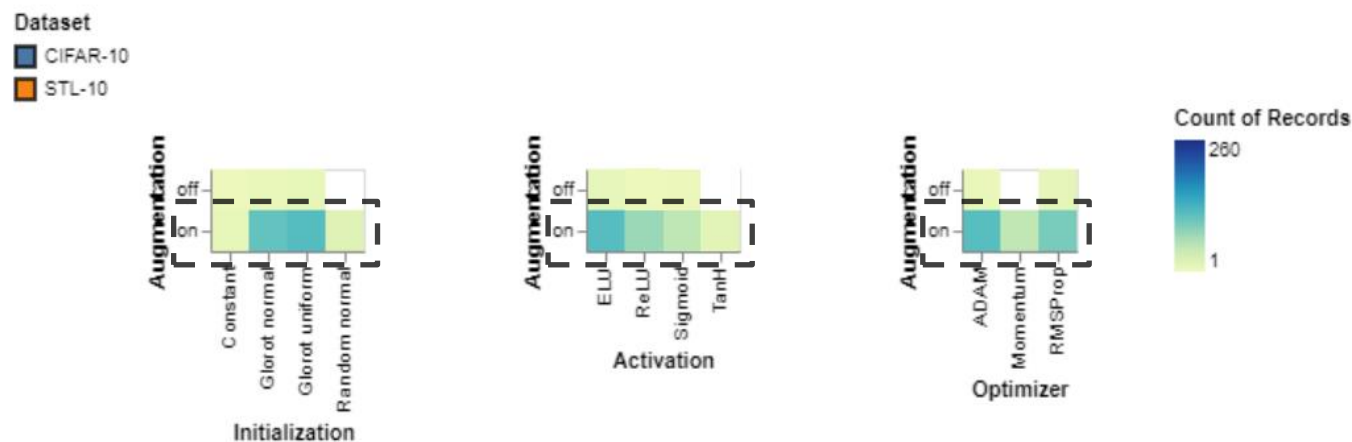
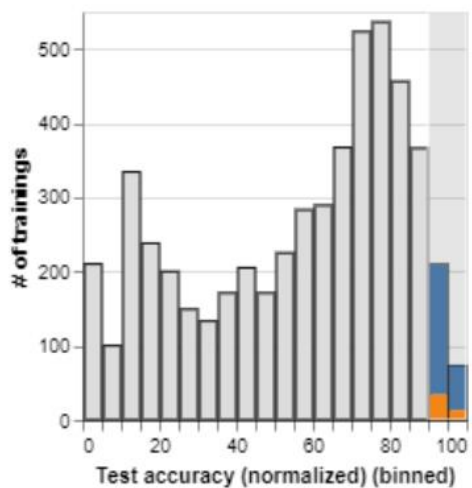
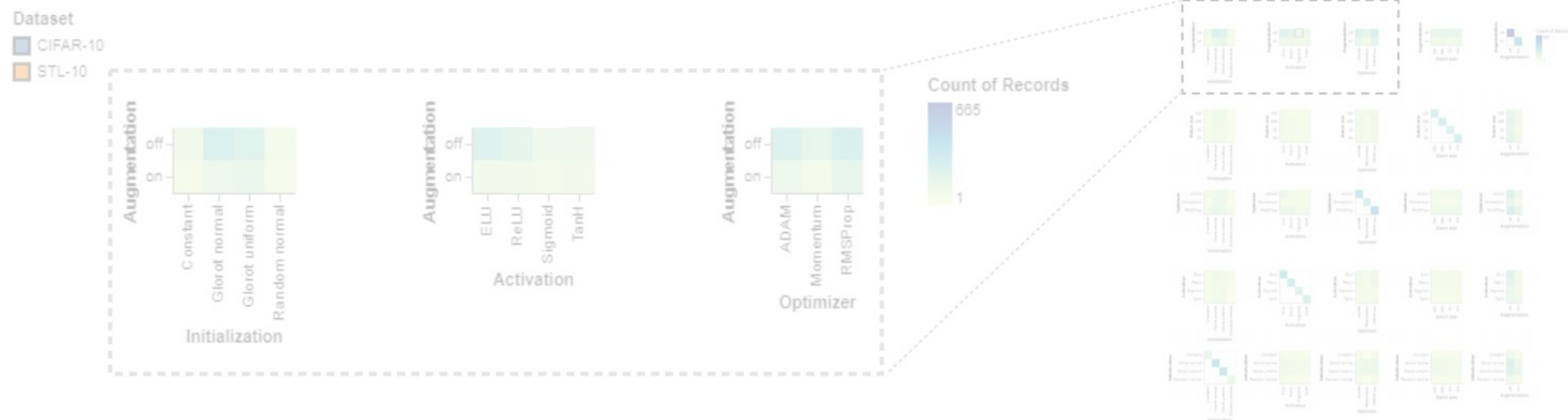
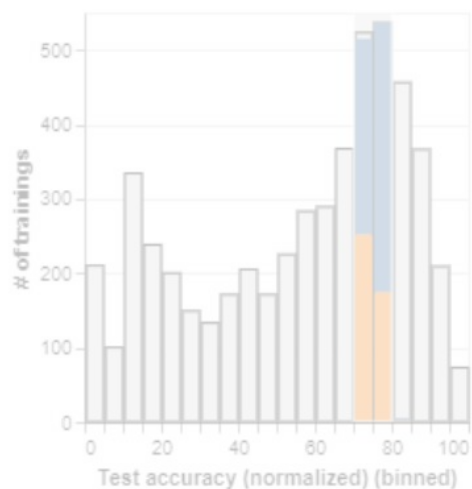
Results



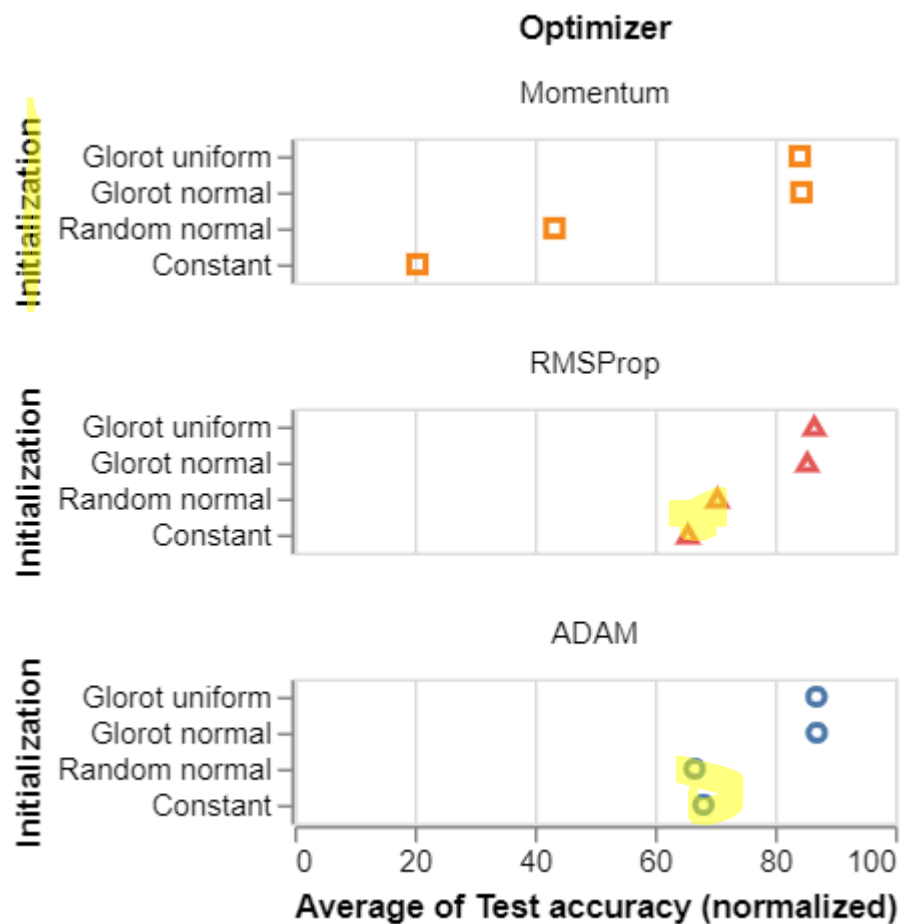
When is augmentation beneficial?



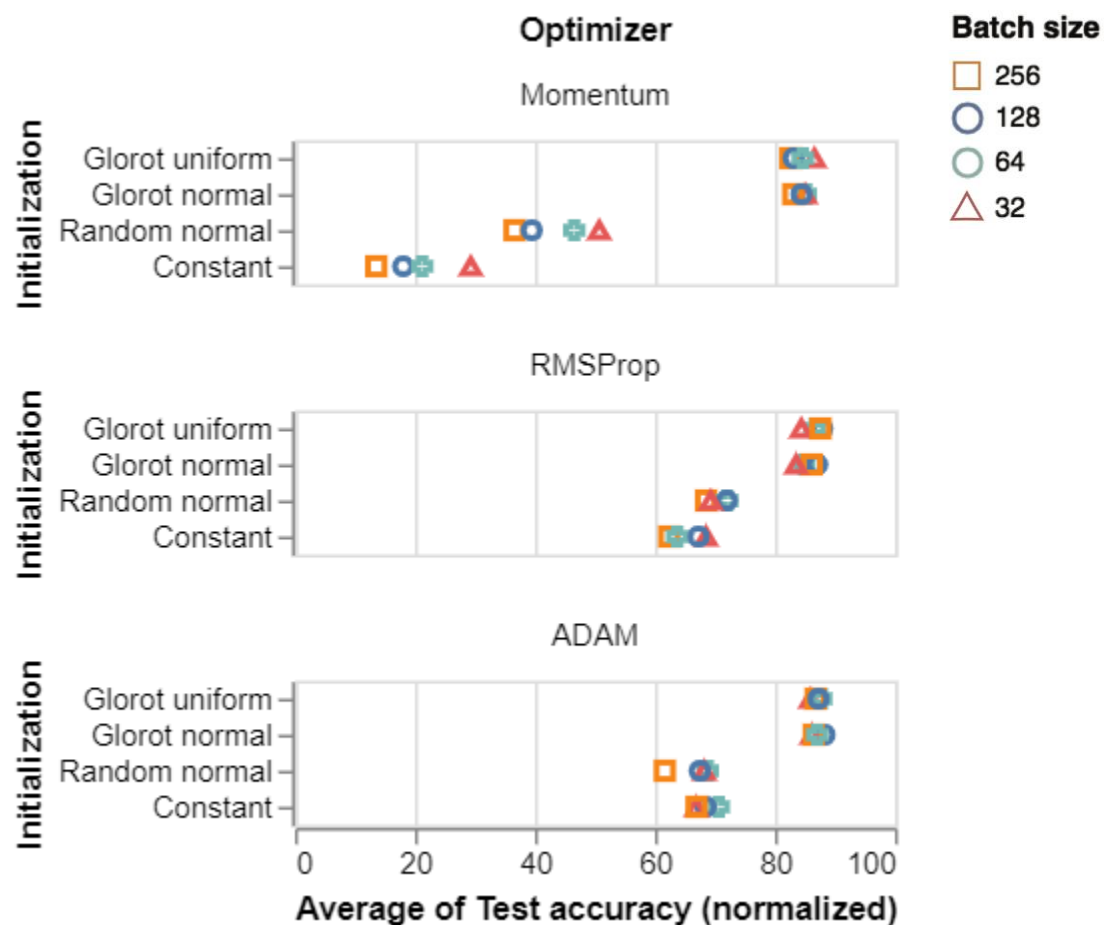
When is augmentation beneficial?



Is optimizer important?



Optimizer – Batch size



Optimizer – Activation



Conclusions

- Augmentation helps for more difficult data sets
- Initialization and activation function are important for top performance

