

# Adventurer: Optimizing Vision Mamba Architecture Designs for Efficiency

## Supplementary Material

### Appendix

#### A. Technical Details

by  $base\_lr * batchsize/512$ ; the color jitter factor is set to 0.3; the mixup alpha and cutmix alpha are set to 0.8 and 1.0, respectively.

Config	Small/Base	Large
input size	128	
optimizer	AdamW	
base learning rate	5e-4	2e-4
weight decay	0.05	0.3
epochs	300	200
optimizer betas	0.9, 0.999	0.9, 0.95
batch size	1024	4096
warmup epochs	5	20
stochastic depth (drop path)	0.1	0.2
layer-wise lr decay	✗	
label smoothing	✗	
random erasing	✗	
Rand Augmentation	✗	
repeated augmentation	✓	
ThreeAugmentation	✓	

Table 8. Settings of Stage One

Config	Small/Base	Large
input size	224	
optimizer	AdamW	
base learning rate	1e-5	2e-5
weight decay	0.1	0.1
epochs	20	50
optimizer betas	0.9, 0.999	0.9, 0.95
batch size	512	512
warmup epochs	5	5
stochastic depth (drop path)	0.4 (S), 0.6 (B)	0.6
layer-wise lr decay	✗	0.95
label smoothing	0.1	
random erasing	✗	
Rand Augmentation	rand-m9-mstd0.5-inc1	
repeated augmentation	✗	
ThreeAugmentation	✗	

Table 10. Settings of Stage Three

Config	Small/Base	Large
input size	224	
optimizer	AdamW	
base learning rate	5e-4	8e-4
weight decay	0.05	0.3
epochs	100	50
optimizer betas	0.9, 0.999	0.9, 0.95
batch size	1024	4096
warmup epochs	5	20
stochastic depth (drop path)	0.2 (S), 0.4 (B)	0.4
layer-wise lr decay	✗	0.9
label smoothing	✗	
random erasing	✗	
Rand Augmentation	✗	
repeated augmentation	✓	
ThreeAugmentation	✓	

Table 9. Settings of Stage Two

We basically follow Mamba-Register’s three-stage training strategy [50] which has been found to be able to effectively prevent Mamba’s over-fitting and save training time. The detailed configurations of each stage are shown in Table 8, Table 9, and Table 10, respectively. For the Tiny-sized model, we directly train with the Small/Base’s stage-2 recipe for 300 epochs since the training time is short and the tiny model is not easy to get overfit. To train the Large-sized model, we make major modifications of the recipe compared with Mamba-Register to further shorten training time. For all stages, the actually learning rate is calculated