,	'. Methods			
		nmasked patche	linear projection + positional embed	dings, ViT
nage → 16×	:16 batches =	huffle		
J	J	masked patches	·	V
		J	mask toker	<u>ns</u> + latent representation
				unshuffle
			shared learned vector	+ positional
				embedding
				decoder (small, shallo
	* decoder	· 只在 pretrain中用到		- (SHUVO) SHUVO
	* skib mu	ask token in encoder and	apply it later in lightweight decoo	der → asummetric
	, ing		Jrg	
1	Tuo			
2	. Tips	V-T-1 -1 -1 -1 -1 -1 -1 -1 -1		
2	. Tips supervised	ViT-L should train with str	ong regularization.	
2	. Tips supervised			
2	. Tips supervised	config optimizer	value AdamW	
2	. Tips supervised	config optimizer base learning rate	value	
2	. Tips supervised	config optimizer base learning rate weight decay optimizer momentum		
2	. Tips supervised	config optimizer base learning rate weight decay optimizer momentum batch size learning rate schedule	value AdamW 1e-4 0.3 $\beta_1, \beta_2 = 0.9, 0.95$ 4096 cosine decay	
2	. Tips supervised	config optimizer base learning rate weight decay optimizer momentum batch size learning rate schedule warmup epochs		
2	. Tips supervised	config optimizer base learning rate weight decay optimizer momentum batch size learning rate schedule warmup epochs training epochs augmentation	$\begin{array}{ c c c c c }\hline & \text{value} & & & & \\ & & \text{AdamW} & & \\ & 1e\text{-}4 & & & \\ & 0.3 & & \\ & \beta_1, \beta_2 = 0.9, 0.95 & \\ & 4096 & & \\ & \text{cosine decay} & & \\ & 20 & & \\ & 300 \text{ (B)}, 200 \text{ (L/H)} & \\ & \text{RandAug } (9, 0.5) \text{ [12]} & & \\ \hline \end{array}$	
2	. Tips supervised	config optimizer base learning rate weight decay optimizer momentum batch size learning rate schedule warmup epochs training epochs augmentation label smoothing [52]	$\begin{array}{ c c c c c }\hline & \text{value} & & & & \\ & & \text{AdamW} & & \\ & 1e\text{-}4 & & & \\ & 0.3 & & \\ & \beta_1, \beta_2 = 0.9, 0.95 & \\ & 4096 & & \\ & \cos & \text{decay} & \\ & 20 & & \\ & 300 \text{ (B)}, 200 \text{ (L/H)} & \\ & \text{RandAug (9, 0.5) [12]} & & \\ & 0.1 & & & \\ \hline \end{array}$	
2	. Tips supervised	config optimizer base learning rate weight decay optimizer momentum batch size learning rate schedule warmup epochs training epochs augmentation label smoothing [52] mixup [69] cutmix [68]	$\begin{array}{ c c c c c }\hline & \text{value} & & & & \\ & & \text{AdamW} & & \\ & 1e\text{-}4 & & & & \\ & 0.3 & & & \\ & \beta_1, \beta_2 \!=\! 0.9, 0.95 & & \\ & 4096 & & & \\ & \cos & \text{decay} & & \\ & 20 & & & \\ & 300 (\text{B}), 200 (\text{L/H}) & & \\ & \text{RandAug} (9, 0.5) [12] & & & \\ & 0.1 & & & & \\ & 0.8 & & & \\ & 1.0 & & & \\ \end{array}$	
2	. Tips supervised	config optimizer base learning rate weight decay optimizer momentum batch size learning rate schedule warmup epochs training epochs augmentation label smoothing [52] mixup [69] cutmix [68] drop path [30]	$\begin{array}{ c c c c c }\hline & \text{value} & & & & \\ & & \text{AdamW} & & \\ & 1e\text{-}4 & & & \\ & 0.3 & & \\ & \beta_1, \beta_2 = 0.9, 0.95 & \\ & 4096 & & \\ & \cos & \text{decay} & \\ & 20 & & \\ & 300 \text{ (B)}, 200 \text{ (L/H)} & \\ & \text{RandAug (9, 0.5) [12]} & & \\ & 0.1 & & \\ & 0.8 & & & \\ \hline \end{array}$	
2	. Tips supervised	config optimizer base learning rate weight decay optimizer momentum batch size learning rate schedule warmup epochs training epochs augmentation label smoothing [52] mixup [69] cutmix [68] drop path [30] exp. moving average (EMA)	value AdamW 1e-4 0.3 β ₁ , β ₂ =0.9, 0.95 4096 cosine decay 20 300 (B), 200 (L/H) RandAug (9, 0.5) [12] 0.1 0.8 1.0 0.1 (B), 0.2 (L/H) 0.9999	
2	. Tips supervised	config optimizer base learning rate weight decay optimizer momentum batch size learning rate schedule warmup epochs training epochs augmentation label smoothing [52] mixup [69] cutmix [68] drop path [30] exp. moving average (EMA)	$\begin{array}{ c c c c c }\hline & \text{value} & & & & \\ & & \text{AdamW} & & \\ & 1e\text{-}4 & & & \\ & 0.3 & & \\ & \beta_1, \beta_2 = 0.9, 0.95 & \\ & 4096 & & \\ & \cos & \text{decay} & \\ & 20 & & \\ & 300 \text{ (B)}, 200 \text{ (L/H)} & \\ & \text{RandAug (9, 0.5) [12]} & & \\ & 0.1 & & \\ & 0.8 & & \\ & 1.0 & & \\ & 0.1 \text{ (B)}, 0.2 \text{ (L/H)} & \\ \hline \end{array}$	
2	. Tips supervised	config optimizer base learning rate weight decay optimizer momentum batch size learning rate schedule warmup epochs training epochs augmentation label smoothing [52] mixup [69] cutmix [68] drop path [30] exp. moving average (EMA)	value AdamW 1e-4 0.3 β ₁ , β ₂ =0.9, 0.95 4096 cosine decay 20 300 (B), 200 (L/H) RandAug (9, 0.5) [12] 0.1 0.8 1.0 0.1 (B), 0.2 (L/H) 0.9999	