

Model	Accuracy (%)	Number Of Trainable parameters
Full Fine-tuning	90	60506624
Soft Prompt From scratch	87	$512 * 10 = 5120$
Soft Prompt Open Delta 10 Token	87	$512 * 10 = 5120$
Soft Prompt Open Delta 1 Token	84.9	$512 * 1 = 512$
Adapter From scratch	90.3	$12 * ((512 + 1) * 8) + ((8+1) * 512) = 104544$
Adapter AdapterHub Bottleneck = 8	90.2	$12 * ((512 + 1) * 8) + ((8+1) * 512) = 104544$
Adapter AdapterHub Bottleneck =1	89.3	$12 * ((512 + 1) * 1) + ((1+1) * 512) = 18444$
Lora From scratch	88.8	$6 * 2 * ((512 + 1) * 8) + ((8+1) * 512) + 6 * 2 * 2 * ((512 + 1) * 8) + ((8+1) * 512) = 313632$
Lora PEFT rank=8	88.5	$6 * 2 * ((512 + 1) * 8) + ((8+1) * 512) + 6 * 2 * 2 * ((512 + 1) * 8) + ((8+1) * 512) = 313632$
Lora PEFT rank=1	88.8	$6 * 2 * ((512 + 1) * 1) + ((1+1) * 512) + 6 * 2 * 2 * ((512 + 1) * 1) + ((1+1) * 512) = 55332$

## Soft Prompt notes:

$n\_token * dim$

Total Parameters =  $5120 + 60506624 = 60511744$

## Adapter notes:

12 layer ( 6 encoder layer + 6 decoder layer)

in\_dim = 512

Bottleneck = 8

out\_dim = 512

Total Parameters =  $104544 + 60506624 = 60611168$

## Lora Notes :

**First Part :** (

6 layer encoder

2 Query and Value

$A\_weights = 512 * 8 + 8$

$B\_weights = 8 * 512 + 512$ )

**Second Part :(**

6 layer Decoder

2 SelfAttention and CrossAttention

2 Query and Value

$A\_weights = 512 * 8 + 8$

$B\_weights = 8 * 512 + 512$ )

Total Parameters =  $60820256$