Table 1: Their Results using ResNet50 backbone.

Method	Backbone	COCO-PS	VIPSeg-VPS	Youtube-VIS-2019
Mask2Former [18]	ResNet50	52.0	-	-
Mask2Former-VIS [16]	ResNet50	-	-	46.4
OMG-Seg	ResNet50	49.9	42.3	46.0
OMG-Seg	ConvNext-L	54.5	50.5	56.2

Table 2: Our Results using ResNet50 backbone.

Method	Backbone	COCO-PS	VIPSeg-VPS	Youtube-VIS-2019
Mask2Former [18]	ResNet50	52.1	-	-
Mask2Former-VIS [16]	ResNet50	-	-	46.4
OMG-Seg	ResNet50	47.9	42.3	45.0
OMG-Seg	ConvNext-L	53.1	48.4	53.4

Table 3: Their Results using ViT backbone.

Backbone	COCO-PS	Youtube-VIS-2019	VIPSeg-VPS
ViT-L (frozen)	34.5	23.2	34.5
ViT-L (learned)	52.2	54.3	48.2
ConvNext-L (frozen)	54.5	56.2	50.5

Table 4: Our Results using ViT backbone.

Backbone	COCO-PS	Youtube-VIS-2019	VIPSeg-VPS
ViT-L (frozen)	31.6	24.2	31.3
ViT-L (learned)	52.2	54.3	45.4
ConvNext-L (frozen)	54.5	56.2	45.8