For the pretraining task we selected extracted backbone of a image segmentation model that was already pretrained on complex tasks. The idea was to start pretraining from a better initialization than just a standard COCO-pretrained resnet50.

model\_backbone\_pretrained = torchvision.models.segmentation.deeplabv3\_resnet50(pretrained=True)

This backbone already has a higher spatial awareness than standard resnet50 with COCO pretraining. After a 12 epoch pretraining on the cityscapes train-extra we have an even better resnet50 for backbone to masked rcnn as we tuned it more to create good feature maps from cityscapes images that include lots of small and detailed objects.

The following Claude 4.0 answers give an idea on how much weights were changed during the pretraining.

## **🎯 Key Findings:**

### **1. Backbone Transfer is Perfect ✅**

* **265 layers transferred** successfully
* **0 missing/unexpected keys** - perfect weight mapping
* **Significant feature differences**: Mean absolute differences of ~1.5-1.8 across all FPN levels

### **2. Your Pretrained Features ARE Different 🎯**

The feature comparison shows substantial differences:

* **Layer 0**: 1.775558 difference
* **Layer 1**: 1.576082 difference
* **Layer 2**: 1.684766 difference
* **Layer 3**: 1.570474 difference