top-1 accuracy drop when Position and Pose vary
(linear eval)

Legend:
- CLIPPretrained
- MAEPretrained
- MLPMixerPretrained1k
- MLPMixerPretrained21k
- ResNet50Pretrained1k
- ResNet50Pretrained21k
- SimCLRPretrained
- ViTPretrained1k
- ViTPretrained21k
- iBotPretrained1k
- iBotPretrained21k

Categories: Background gap, Size gap, Lighting gap, Pose gap, Position gap