Appendix

A. Architecture

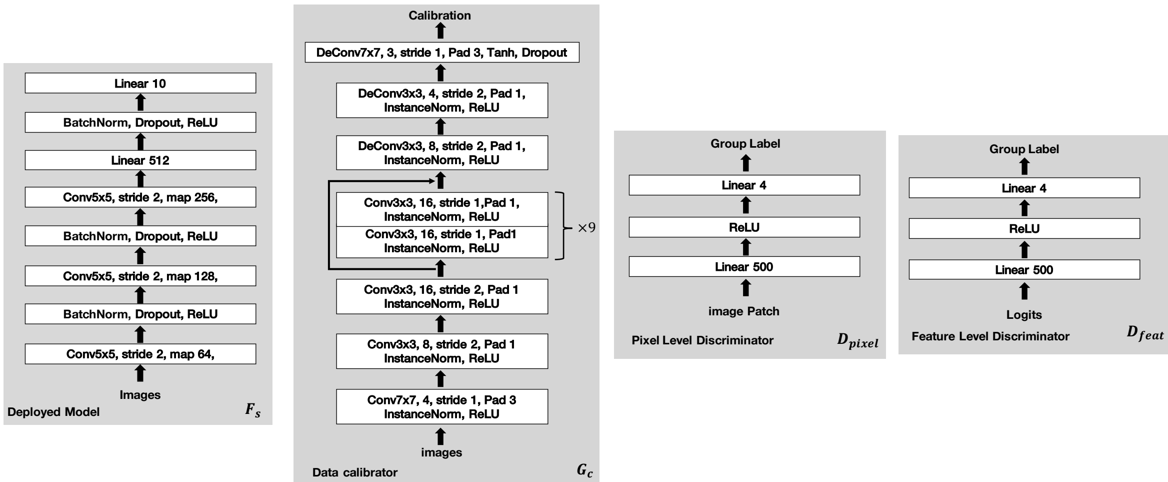

Figure 7. **Network architectures used for digits experiments** . We show the source classifier F_s, proposed calibrator G_c, pixel level domain discriminator D_{pixel} and feature level domain discriminator D_{feat}.

B. L_∞ norm of the calibration

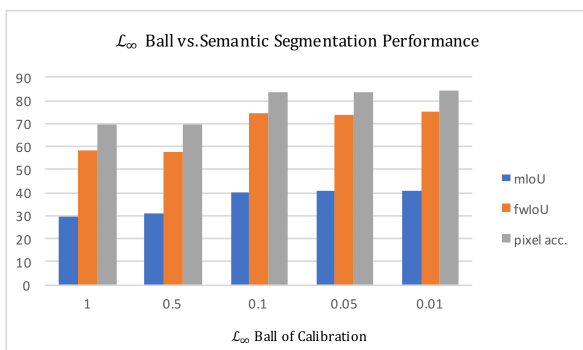

Figure 8. **Performance vs. L_∞ norm of calibration produced by the calibrator**. Constraining L_∞ norm of the calibration has a positive impact on the overall performance.

Suggested by Figure 8, constraining the L_∞ norm of the calibration to a level that is not perceivable work even better compared to larger L_∞ norm counterparts. Our hypothesis is that: (1) without L_∞ norm constrain, the semantic consistency of the inputs might be hurt through the calibration. (2) The fact that small L_∞ norm works better are likely due to the fact that calibration manipulates non-robust features, which might not be obvious to our eyes.

C. Overhead relative to main model

GTA5 to CityScapes	N. of Param.(M)	Flops(G)
DRN-26	20.6	200
Data Calibrator	0.05	2.67
Digits	N. of Param.(M)	Flops(G)
LeNet	3.13	0.03
Data Calibrator	0.18	0.02

Table 3. **Overhead of data calibrator**. We show that our calibrator is light-weight both in terms of number of parameters and flops. Even for network as tiny as LeNet, the calibrator is small compared to it