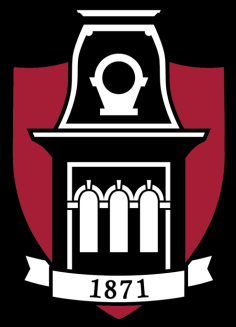


Carnegie Mellon University

JUNE 18-22, 2023

CVPR

VANCOUVER, CANADA

FREDOM: Fairness Domain Adaptation Approach to Semantic Scene Understanding

Thanh-Dat Truong[1], Ngan Le[1], Bhiksha Raj[2], Jackson Cothren[3], Khoa Luu[1]

[1]CVIU Lab, University of Arkansas

[2]Carnegie Mellon University

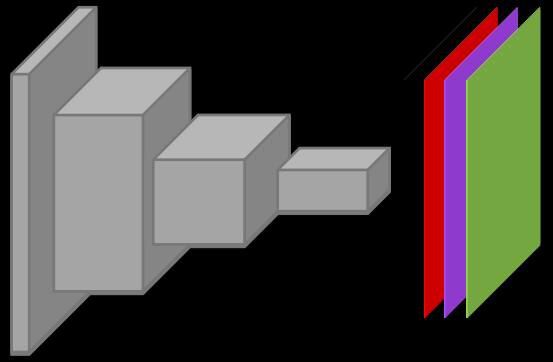
[3]Department of Geosciences, University of Arkansas

https://uark-cviu.github.io/

Unsupervised Domain Adaptation

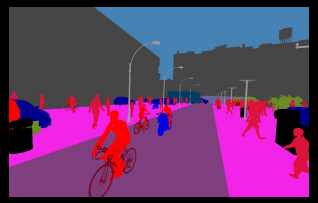
Source Domain

Supervised Learning



Prediction

Ground-Truth



Target Domain

Domain Adaptation

Prediction

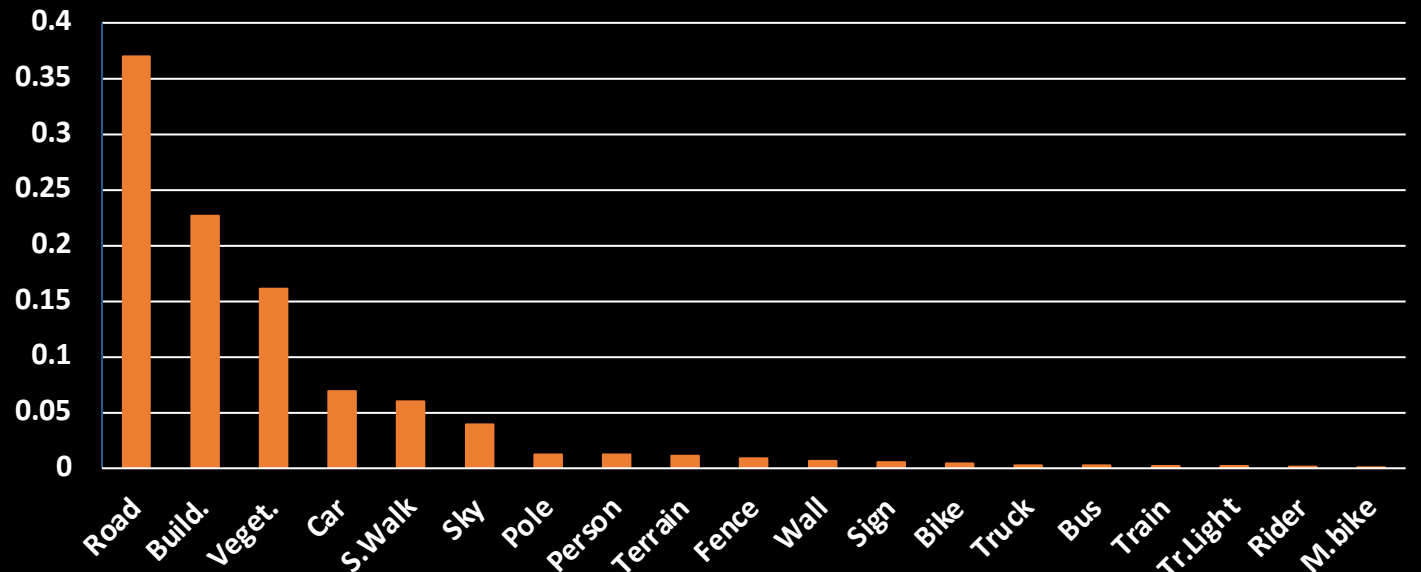
Ground-Truth

Not Available

Fairness Problem in Domain Adaptation

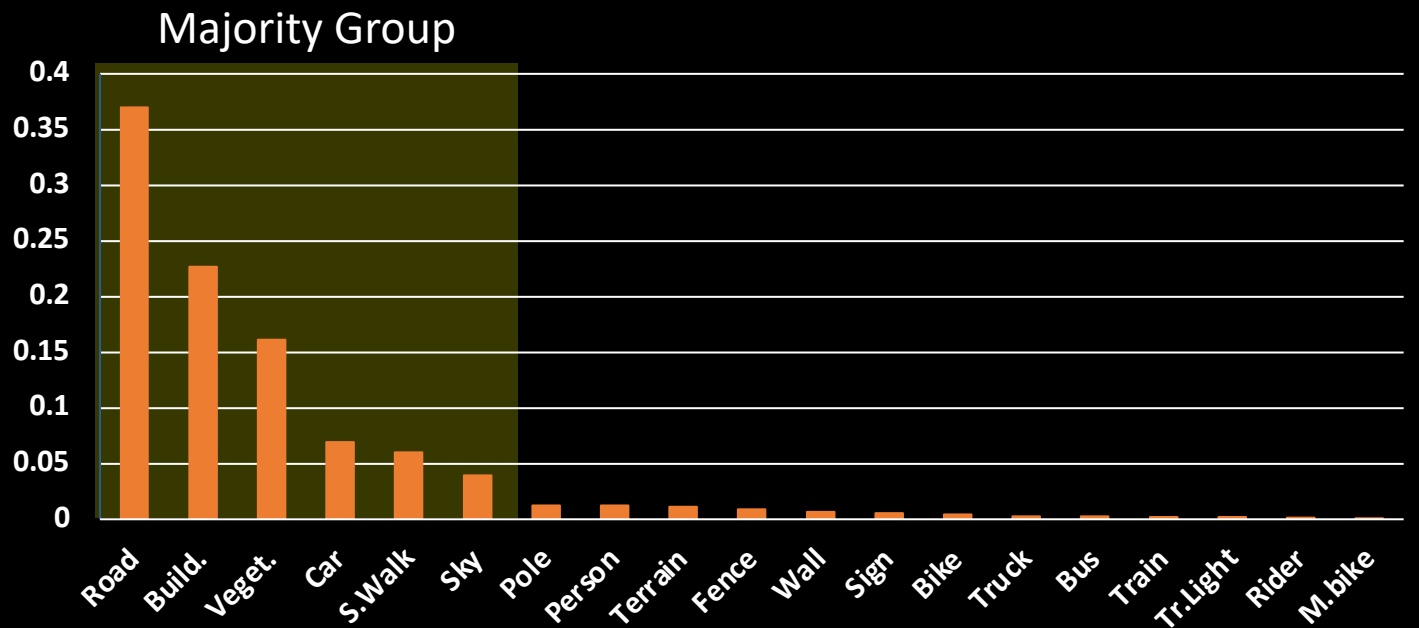
Fairness Problem in Domain Adaptation

The Performance of Segmentation Models on Majority and Minority Groups of Classes



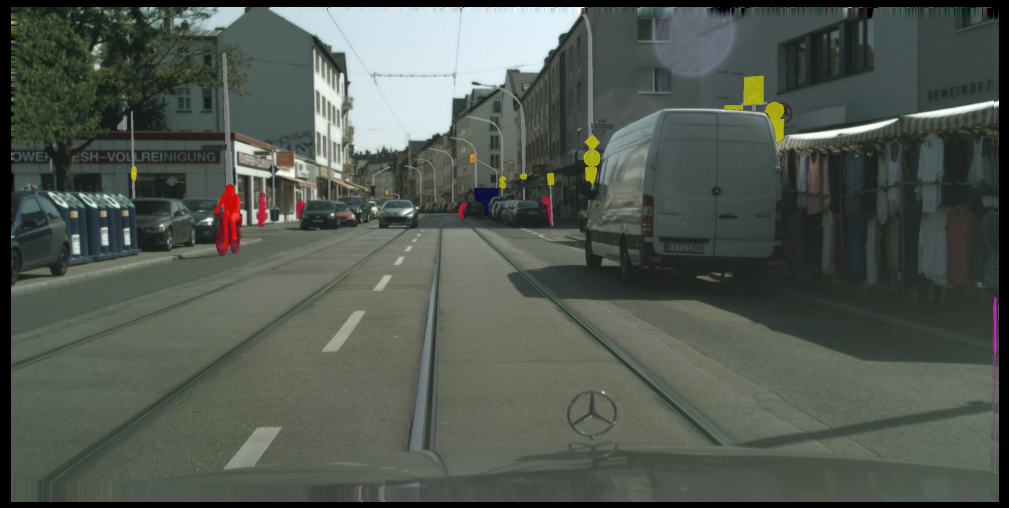
Fairness Problem in Domain Adaptation

The Performance of Segmentation Models on Majority and Minority Groups of Classes

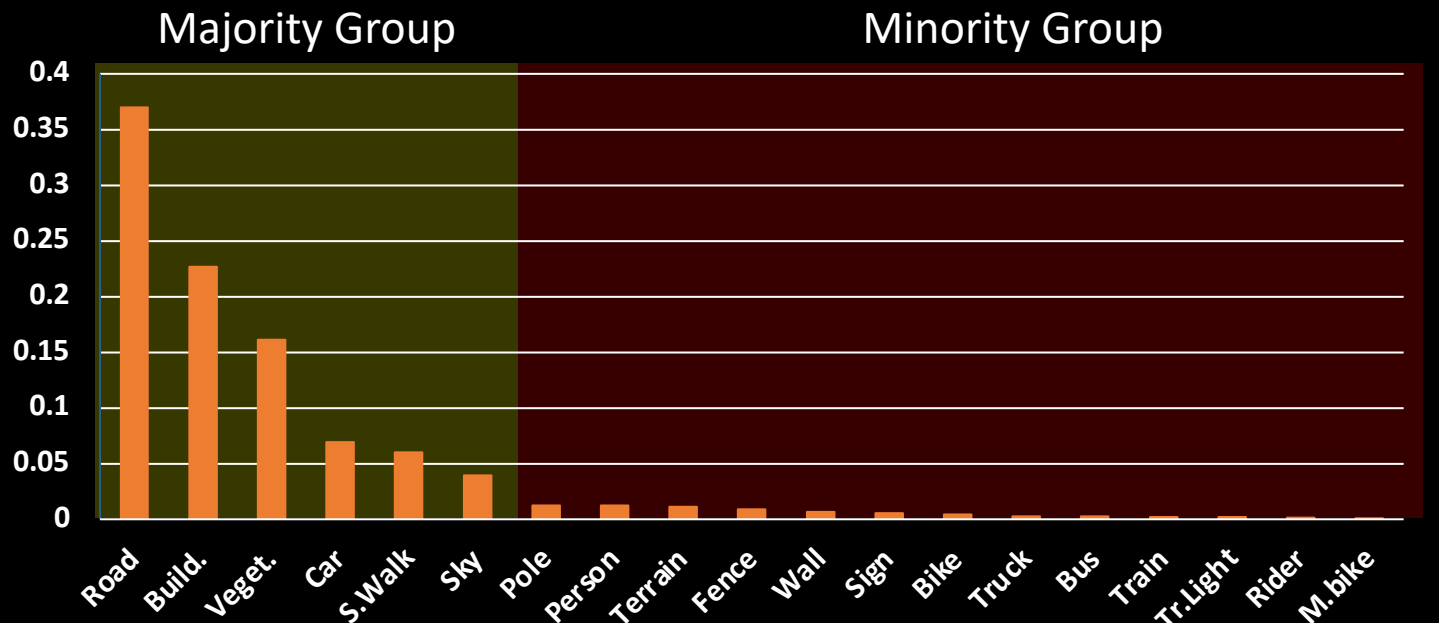


Containing Many Pixels

Fairness Problem in Domain Adaptation



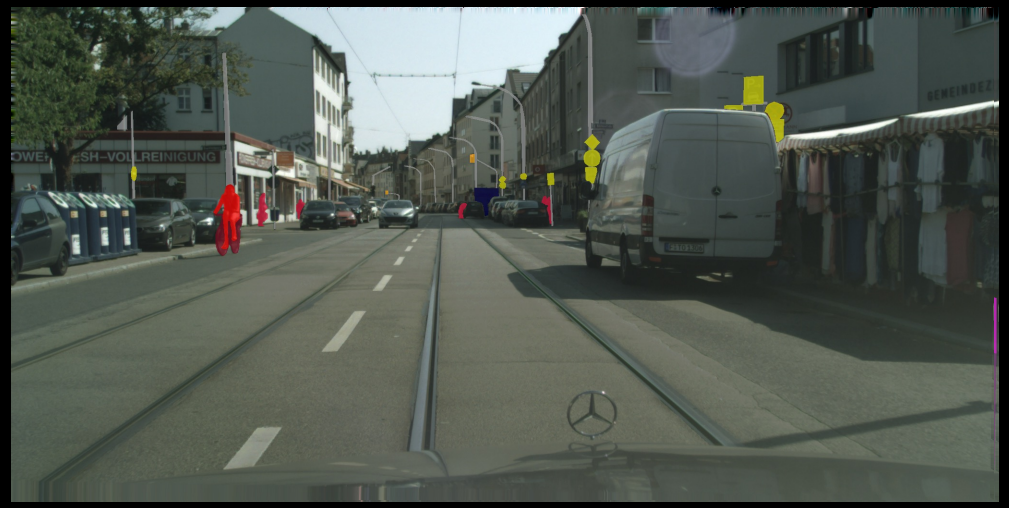
The Performance of Segmentation Models on Majority and Minority Groups of Classes



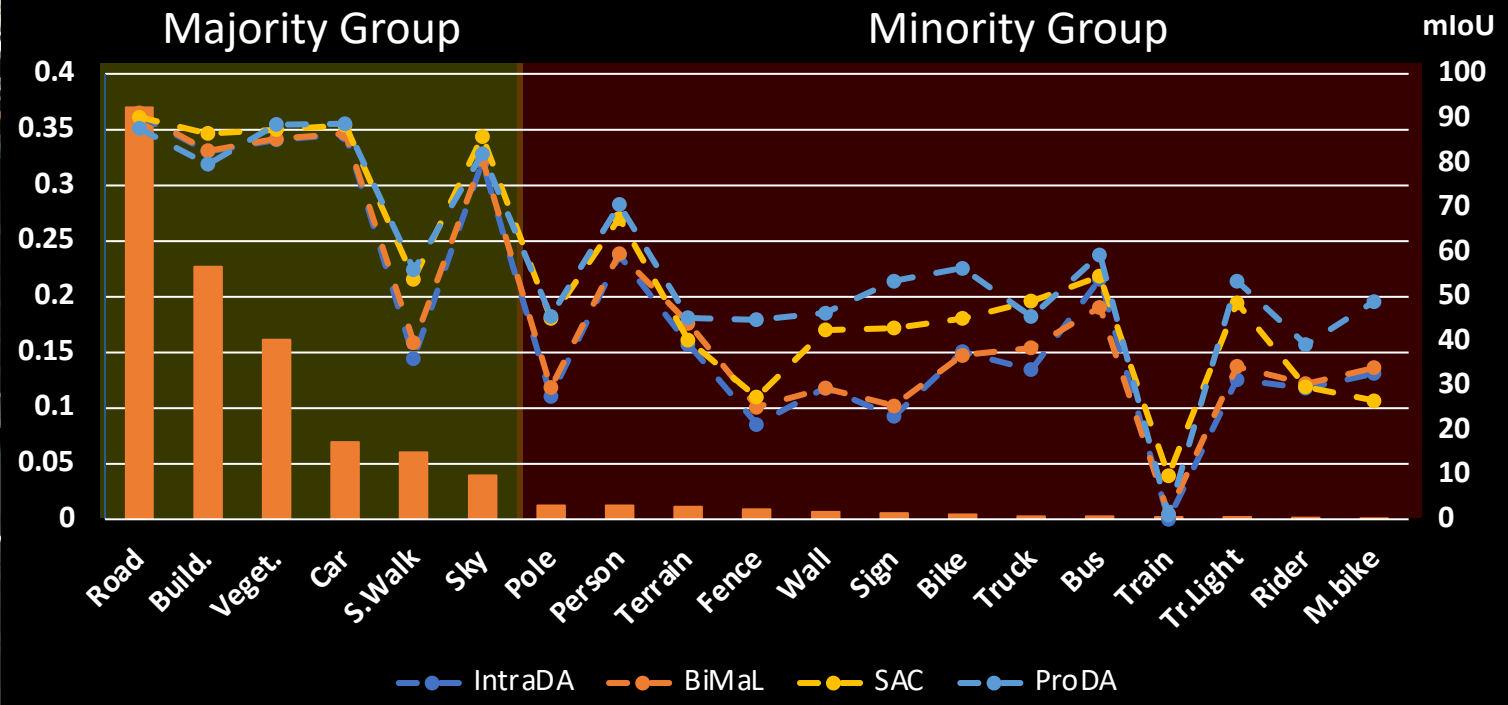
Containing Many Pixels

Containing Less Pixels

Fairness Problem in Domain Adaptation



The Performance of Segmentation Models on Majority and Minority Groups of Classes

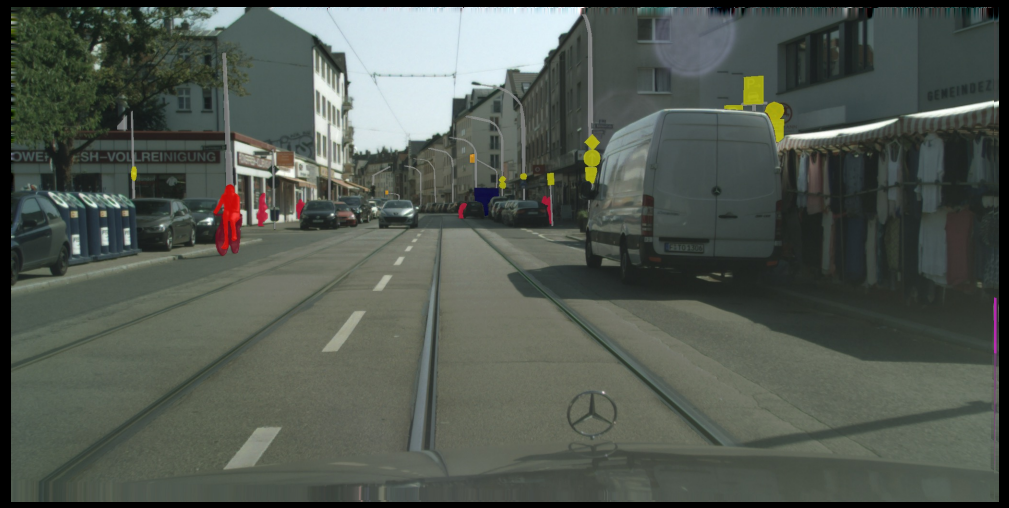


Containing Many Pixels

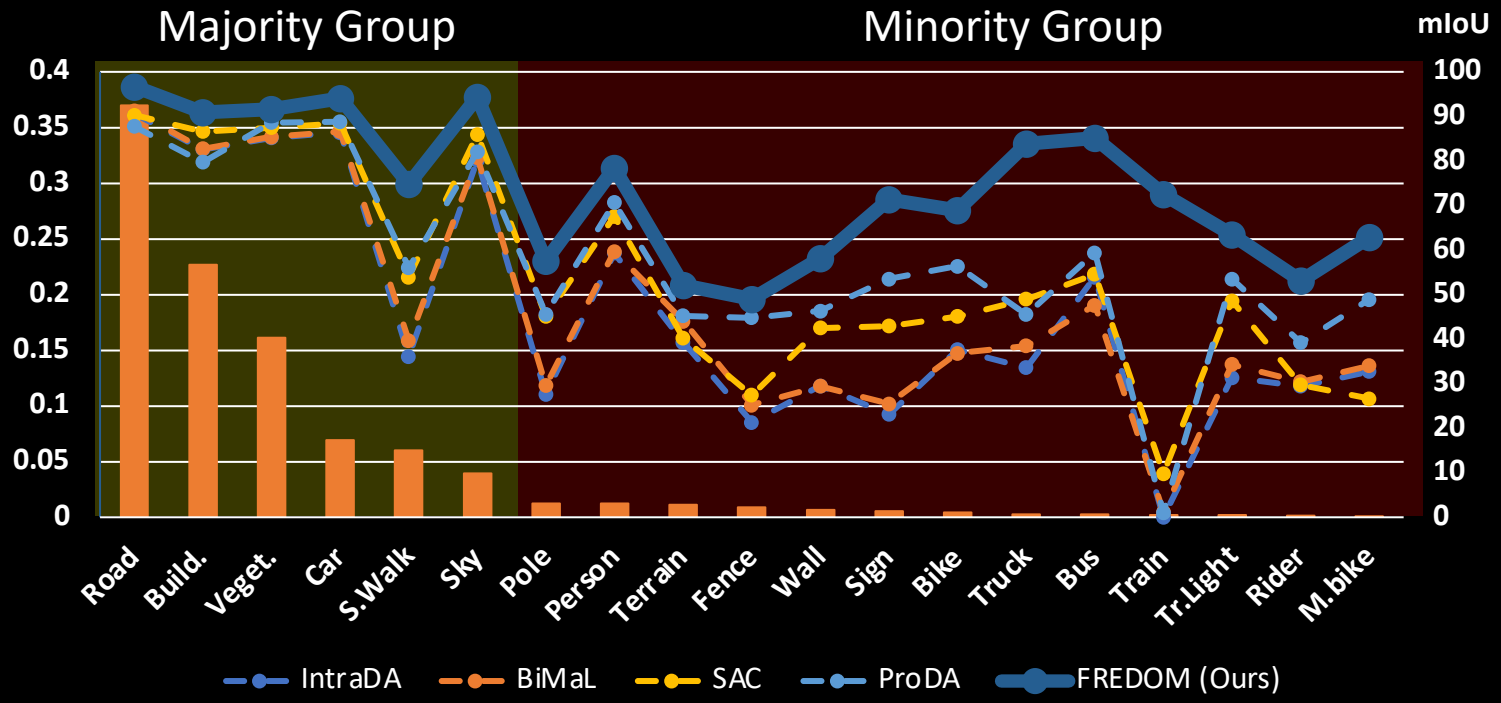
Containing Less Pixels

Low Performance on Minority Group

Fairness Problem in Domain Adaptation



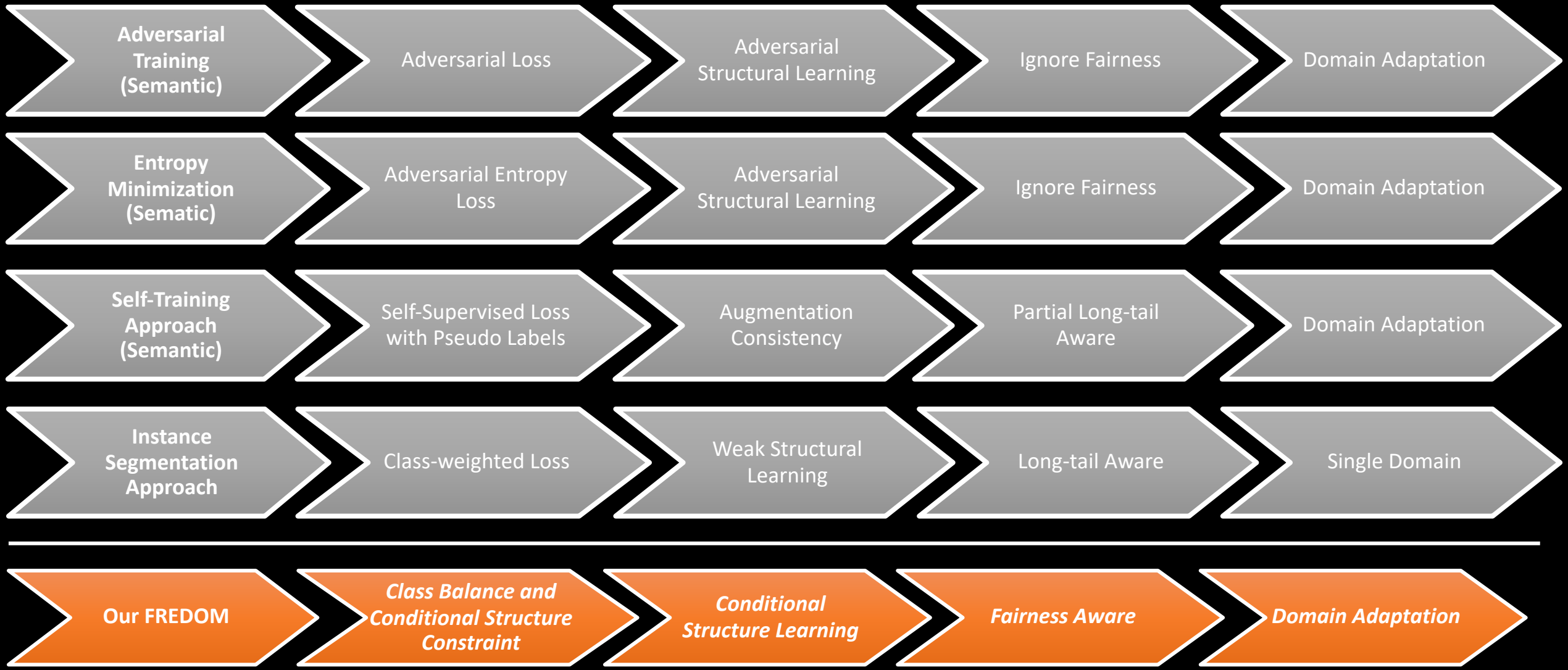
The Performance of Segmentation Models on Majority and Minority Groups of Classes



Containing Many Pixels *Containing Less Pixels*

Our FREDOM approach improves performance of the minority group to promote fairness between classes

Motivation


Adversarial Training (Semantic)
Adversarial Loss
Adversarial Structural Learning
Ignore Fairness
Domain Adaptation
Entropy Minimization (Sematic)
Adversarial Entropy Loss
Adversarial Structural Learning
Ignore Fairness
Domain Adaptation
Self-Training Approach (Semantic)
Self-Supervised Loss with Pseudo Labels
Augmentation Consistency
Partial Long-tail Aware
Domain Adaptation
Instance Segmentation Approach
Class-weighted Loss
Weak Structural Learning
Long-tail Aware
Single Domain
Our FREDOM
Class Balance and Conditional Structure Constraint
Conditional Structure Learning
Fairness Aware
Domain Adaptation

Contributions

- Present a novel Fairness metric between classes for semantic segmentation
- Propose new Fairness Domain Adaptation approach to Semantic Segmentation
  - Promote fairness by a new fairness treatment loss from class distributions
  - Impose consistency of segmentation maps by a novel Conditional Structural Constraint
  - Model Conditional Structural Constraint by the Conditional Structure Network
- Achieve State-of-the-Art Performance on Domain Adaptation Benchmarks and Promote Fairness of the model predictions

Fairness Objective

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{c_i c_j} \underbrace{\left| \mathbb{E}_{x \in X} \sum_k \mathcal{L}(y^k = c_i) - \mathbb{E}_{x \in X} \sum_k \mathcal{L}(y^k = c_j) \right|}$$

Minimize the Difference of Error Rates Between Classes
So That the Model Behaves Fairly Between Classes

Fairness Objective

$$\sum_{c_i c_j} \left| \mathbb{E}_{x \in \mathcal{X}} \sum_k \mathcal{L}(y^k = c_i) - \mathbb{E}_{x \in \mathcal{X}} \sum_k \mathcal{L}(y^k = c_j) \right| \leq \underbrace{2C\left[\mathbb{E}_{x_s, \hat{y}_s \sim q_s(x_s, \hat{y}_s)} \mathcal{L}_s(y_s, \hat{y}_s) + \mathbb{E}_{x_t \sim p_t(x_t)} \mathcal{L}_t(y_t)\right]}$$

Standard Unsupervised Domain Adaptation

Fairness Objective

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \left[\mathbb{E}_{x_s, \hat{y}_s \sim q_s(x_s, \hat{y}_s)} \mathcal{L}_s(y_s, \hat{y}_s) + \mathbb{E}_{x_t \sim p_t(x_t)} \mathcal{L}_t(y_t) \right]$$
$$= argmin_\theta \int \mathcal{L}_s(y_s, \hat{y}_s) q_s(y_s, \hat{y}_s) dy_s d\hat{y}_s \; + \; \int \mathcal{L}_t(y_t) p_t(y_t) dy_t$$

Why Does The UDA Model Behave Unfairly?

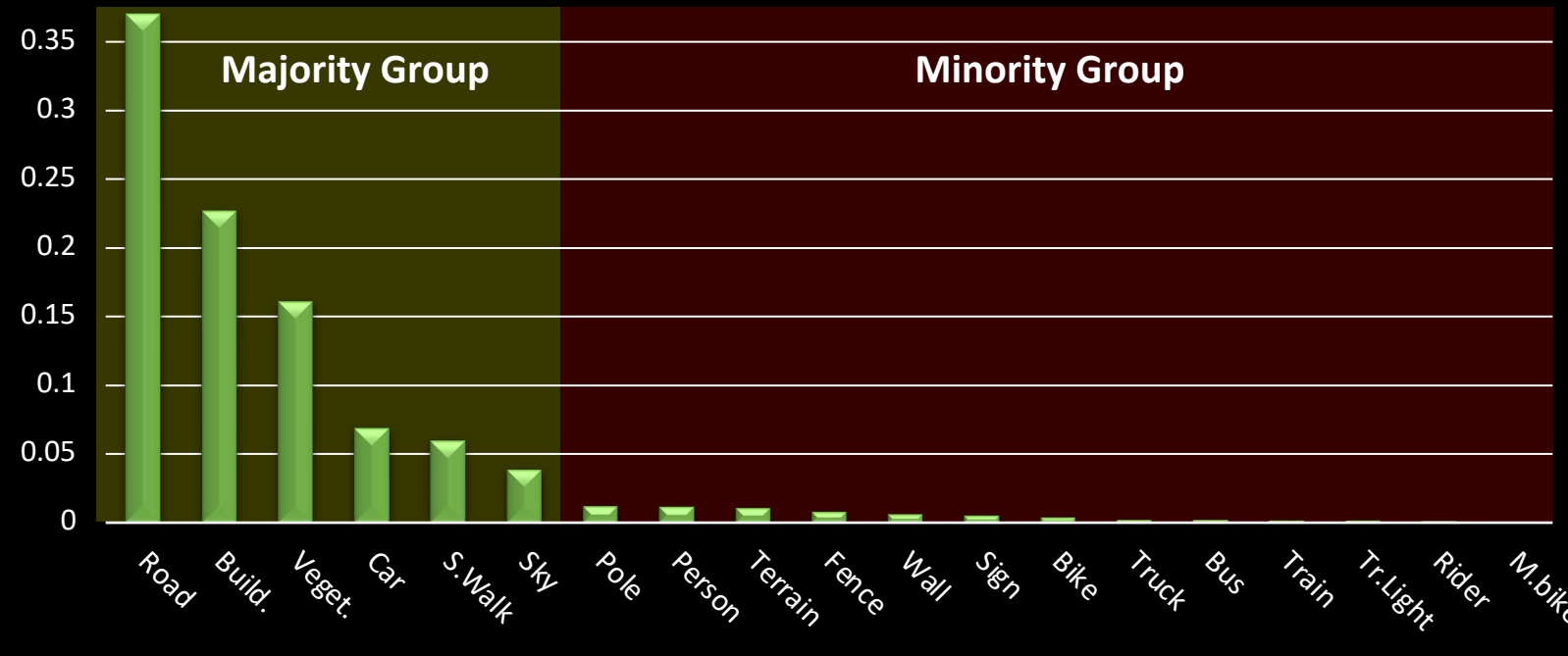
$$\theta^* = argmin_\theta \int \mathcal{L}_s(y_s, \hat{y}_s) q_s(y_s, \hat{y}_s) dy_s d\hat{y}_s + \int \mathcal{L}_t(y_t) p_t(y_t) dy_t$$

$$= argmin_\theta \int \sum_{k=1}^{N} \mathcal{L}_s(y_s^k, \hat{y}_s^k) q_s(y_s^k) q\_s\left(y_s^{\setminus k} \middle| y_s^k\right) q_s(\hat{y}_s|y_s) dy_s d\hat{y}_s + \int \sum_{k=1}^{N} \mathcal{L}_t(y_t) p_t(y_t^k) p_t(y_t^{\setminus k}) dy_t$$

Suffer Imbalance Distributions

The Class Distribution based on the Number of Pixels

Gradients Updated to Predictions of Classes in the Majority Group Largely Dominant the ones in the Minority Group



The Proposed Fairness Approach

$$\theta^* = argmin_\theta \mathbb{E}_{x_s,\hat{y}_s \sim q_s(y_s,\hat{y}_s)} \mathcal{L}_s(y_s,\hat{y}_s) \frac{q'_s(y_s,\hat{y}_s)}{q_s(y_s,\hat{y}_s)} + \mathbb{E}_{x_t \sim p_t(x_t)} \mathcal{L}_t(y_t) \frac{p'_t(y_t)}{p_t(y_t)}$$

Ideal Distributions
Where the Learned Model Behave Fairly

$$\theta^* \cong argmin_\theta \mathbb{E}_{y_s,\hat{y}_s \sim q_s(y_s,\hat{y}_s)} \left[\mathcal{L}_s(y_s,\hat{y}_s) + \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{q'_s(y_s^i)}{q_s(y_s^i)} + \log \frac{q'_s\left(y_s^{\setminus i} | y_s^i\right)}{q_s\left(y_s^{\setminus i} | y_s^i\right)} \right) \right] + \mathbb{E}_{y_t \sim p_t(y_t)} \left[\mathcal{L}_t(y_t) + \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{p'_t(y_t^i)}{q_s(y_s^i)} + \log \frac{q'_t\left(y_t^{\setminus i} | y_t^i\right)}{q_t\left(y_t^{\setminus i} | y_t^i\right)} \right) \right]$$

$$\leq argmin_\theta \mathbb{E}_{y_s,\hat{y}_s \sim q_s(y_s,\hat{y}_s)} \left[\mathcal{L}_s(y_s,\hat{y}_s) + \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{q'_s(y_s^i)}{q_s(y_s^i)} - \log q_s\left(y_s^{\setminus i} | y_s^i\right) \right) \right] + \mathbb{E}_{y_t \sim p_t(y_t)} \left[\mathcal{L}_t(y_t) + \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{p'_t(y_t^i)}{q_s(y_s^i)} - \log q_t\left(y_t^{\setminus i} | y_t^i\right) \right) \right]$$

The Proposed Fairness Approach

$$\theta^* = argmin_\theta \mathbb{E}_{x_s,\hat{y}_s \sim q_s(y_s,\hat{y}_s)} \mathcal{L}_s(y_s,\hat{y}_s) \frac{q'_s(y_s,\hat{y}_s)}{q_s(y_s,\hat{y}_s)} + \mathbb{E}_{x_t \sim p_t(x_t)} \mathcal{L}_t(y_t) \frac{p'_t(y_t)}{p_t(y_t)}$$

Ideal Distributions
Where the Learned Model Behave Fairly

$$\theta^* \cong argmin_\theta \mathbb{E}_{y_s,\hat{y}_s \sim q_s(y_s,\hat{y}_s)} \left[\mathcal{L}_s(y_s,\hat{y}_s) + \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{q'_s(y_s^i)}{q_s(y_s^i)} + \log \frac{q'_s\left(y_s^{\setminus i} | y_s^i\right)}{q_s\left(y_s^{\setminus i} | y_s^i\right)} \right) \right] + \mathbb{E}_{y_t \sim p_t(y_t)} \left[\mathcal{L}_t(y_t) + \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{p'_t(y_t^i)}{q_s(y_s^i)} + \log \frac{q'_t\left(y_t^{\setminus i} | y_t^i\right)}{q_t\left(y_t^{\setminus i} | y_t^i\right)} \right) \right]$$

$$\leq argmin_\theta \mathbb{E}_{y_s,\hat{y}_s \sim q_s(y_s,\hat{y}_s)} \left[\mathcal{L}_s(y_s,\hat{y}_s) + \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{q'_s(y_s^i)}{q_s(y_s^i)} - \log q_s\left(y_s^{\setminus i} | y_s^i\right) \right) \right] + \mathbb{E}_{y_t \sim p_t(y_t)} \left[\mathcal{L}_t(y_t) + \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{p'_t(y_t^i)}{q_s(y_s^i)} - \log q_t\left(y_t^{\setminus i} | y_t^i\right) \right) \right]$$

Standard Domain Adaptation Loss

The Proposed Fairness Approach

$$\theta^* = argmin_\theta \mathbb{E}_{x_s,\hat{y}_s \sim q_s(y_s,\hat{y}_s)} \mathcal{L}_s(y_s,\hat{y}_s) \frac{q'_s(y_s,\hat{y}_s)}{q_s(y_s,\hat{y}_s)} + \mathbb{E}_{x_t \sim p_t(x_t)} \mathcal{L}_t(y_t) \frac{p'_t(y_t)}{p_t(y_t)}$$

Ideal Distributions
Where the Learned Model Behave Fairly

$$\theta^* \cong argmin_\theta \mathbb{E}_{y_s,\hat{y}_s \sim q_s(y_s,\hat{y}_s)} \left[\mathcal{L}_s(y_s,\hat{y}_s) + \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{q'_s(y_s^i)}{q_s(y_s^i)} + \log \frac{q'_s(y_s^{\setminus i}|y_s^i)}{q_s(y_s^{\setminus i}|y_s^i)} \right) \right] + \mathbb{E}_{y_t \sim p_t(y_t)} \left[\mathcal{L}_t(y_t) + \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{p'_t(y_t^i)}{q_s(y_s^i)} + \log \frac{q'_t(y_t^{\setminus i}|y_t^i)}{q_t(y_t^{\setminus i}|y_t^i)} \right) \right]$$

$$\leq argmin_\theta \mathbb{E}_{y_s,\hat{y}_s \sim q_s(y_s,\hat{y}_s)} \left[\mathcal{L}_s(y_s,\hat{y}_s) + \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{q'_s(y_s^i)}{q_s(y_s^i)} - \log q_s\left(y_s^{\setminus i}|y_s^i\right) \right) \right] + \mathbb{E}_{y_t \sim p_t(y_t)} \left[\mathcal{L}_t(y_t) + \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{p'_t(y_t^i)}{q_s(y_s^i)} - \log q_t\left(y_t^{\setminus i}|y_t^i\right) \right) \right]$$

Minority Group
e.g., Tr.Light, Sign, Person

Majority Group
e.g., Car, Sky, Sidewalk

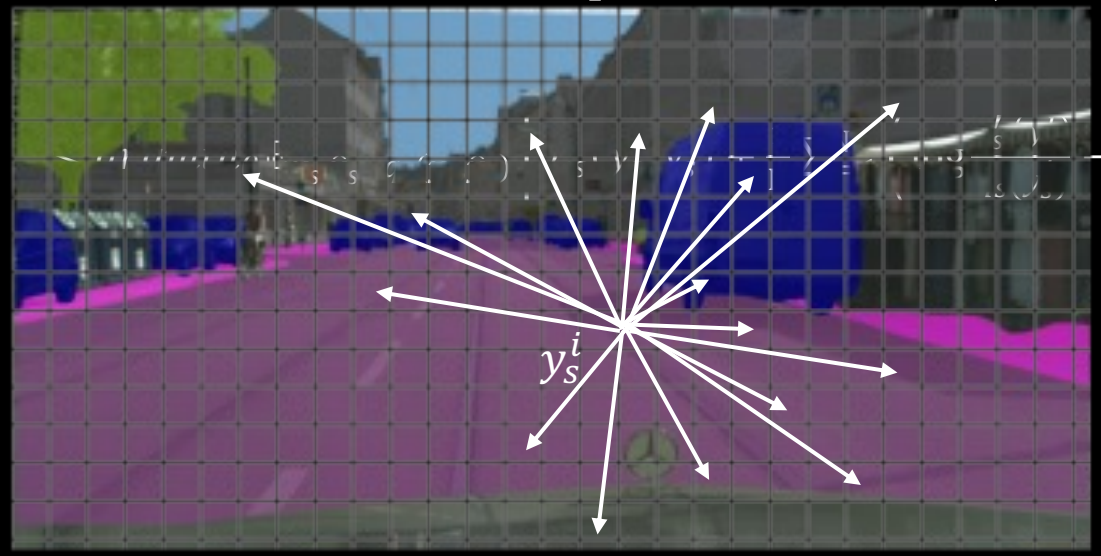
Fairness Treatment Loss
From Class Distribution

The Proposed Fairness Approach

$$\theta^* = argmin_\theta \mathbb{E}_{x_s,\hat{y}_s \sim q_s(y_s,\hat{y}_s)} \mathcal{L}_s(y_s,\hat{y}_s) \frac{q'_s(y_s,\hat{y}_s)}{q_s(y_s,\hat{y}_s)} + \mathbb{E}_{x_t \sim p_t(x_t)} \mathcal{L}_t(y_t) \frac{p'_t(y_t)}{p_t(y_t)}$$

Ideal Distributions
Where the Learned Model Behave Fairly

$$\theta^* \cong argmin_\theta \mathbb{E}_{y_s,\hat{y}_s \sim q_s(y_s,\hat{y}_s)} \left[\mathcal{L}_s(y_s,\hat{y}_s) + \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{q'_s(y_s^i)}{q_s(y_s^i)} + \log \frac{q'_s\left(y_s^{\setminus i} | y_s^i\right)}{q_s\left(y_s^{\setminus i} | y_s^i\right)} \right) \right] + \mathbb{E}_{y_t \sim p_t(y_t)} \left[\mathcal{L}_t(y_t) + \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{p'_t(y_t^i)}{q_s(y_s^i)} + \log \frac{q'_t\left(y_t^{\setminus i} | y_t^i\right)}{q_t\left(y_t^{\setminus i} | y_t^i\right)} \right) \right]$$



$$\left. + \log q_s\left(y_s^{\setminus i} | y_s^i\right) \right) \Big] + \mathbb{E}_{y_t \sim p_t(y_t)} \left[\mathcal{L}_t(y_t) + \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\log \frac{p'_t(y_t^i)}{q_s(y_s^i)} - \log q_t\left(y_t^{\setminus i} | y_t^i\right) \right) \right]$$

Conditional Structural Constraint

Conditional Structure Network

$$\Theta^* = -argmin_\Theta \mathbb{E}_{y_s \sim \mathcal{Y}_s, \sigma^k \sim \prod} \log q_s\left(y_s^{\backslash i} | y_s^k\right)$$

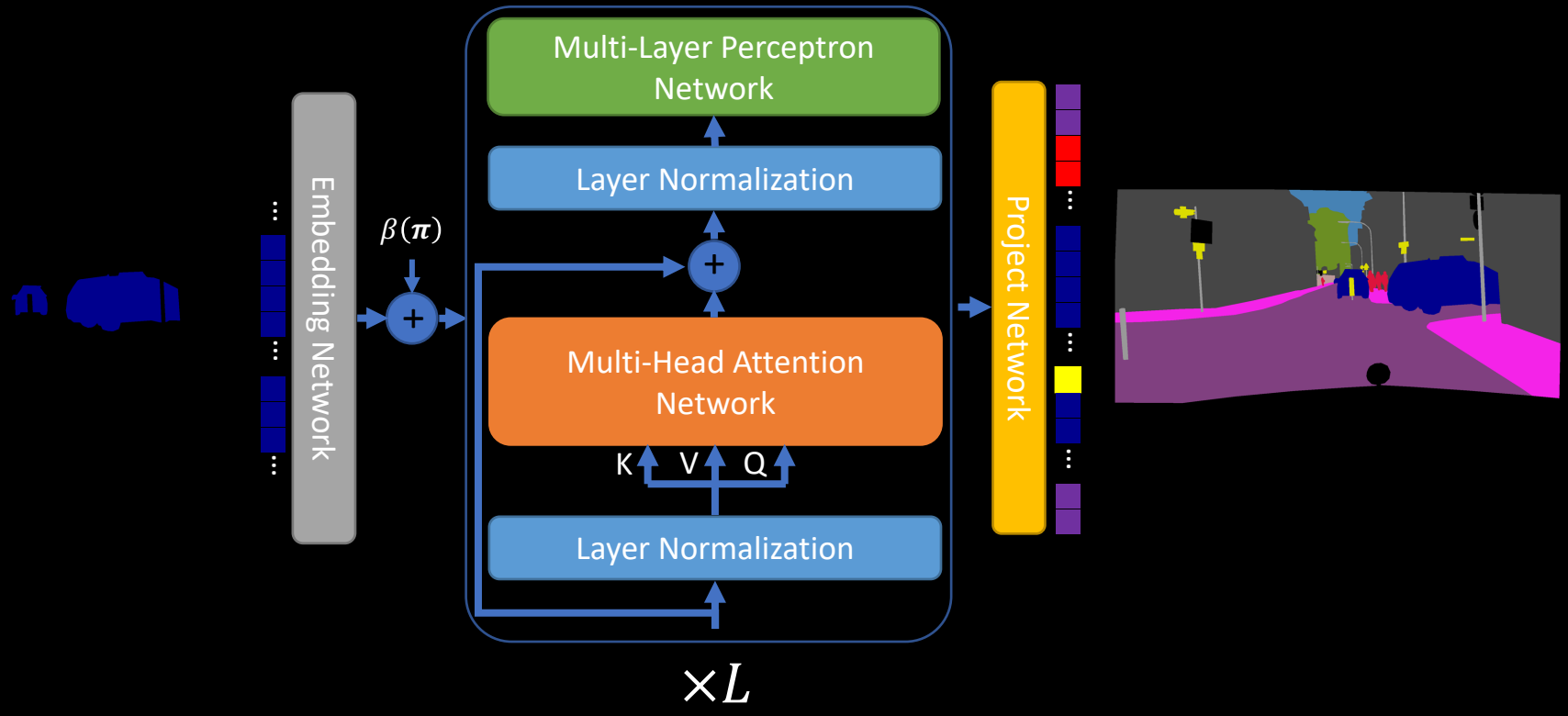
$$= \underbrace{-argmin_\Theta \mathbb{E}_{y_s \sim \mathcal{Y}_s, \sigma^k \sim \prod} \sum_{i=1}^{N-1} \log q_s\left(y_s^{\sigma_i^k} | y_s^{\sigma_1^k} \dots y_s^{\pi_{i-1}^k} y_s^k\right)}$$

Solving by Pixel RNN (or Pixel CNN) is ineffective when N is a large number

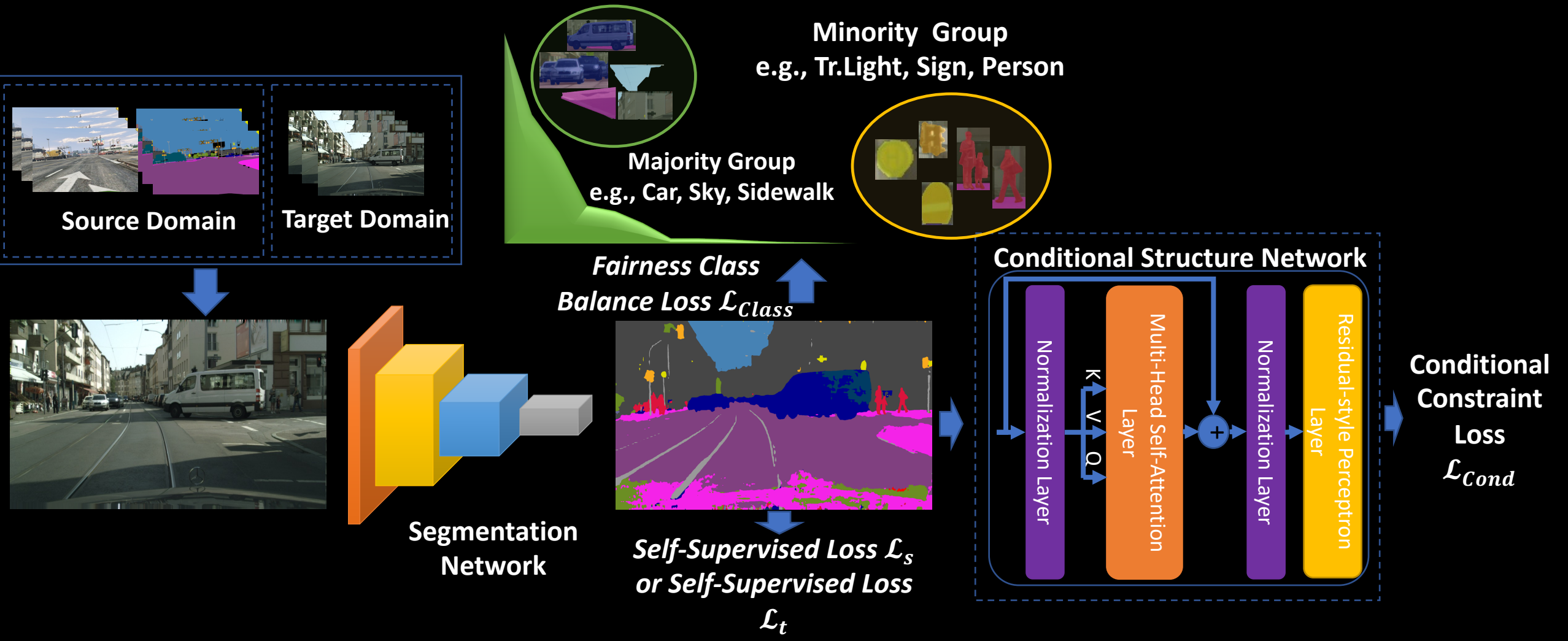
$$\Theta^* = \underbrace{-argmin_\Theta \mathbb{E}_{y_s \sim \mathcal{Y}_s, m \sim \mathcal{M}} \log q_s(y_s \odot (1-m) \mid y_s \odot m))}$$ where m is the conditional mask

Learn the Conditional Structure Constraints By the Multi-head Self-Attention Network

Conditional Structure Network



The Proposed FREDOM Framework



Thank You For Watching