Towards Unsupervised Domain Bridging via Image Degradation in Semantic Segmentation

Wangkai Li[1], Rui Sun[1], Huayu Mai[1], Tianzhu Zhang[1, 2]

[1]University of Science and Technology of China, [2]Deep Space Exploration Laboratory

TLDR. Bridging domains via controllable image degradation and time-aware semantic compensation

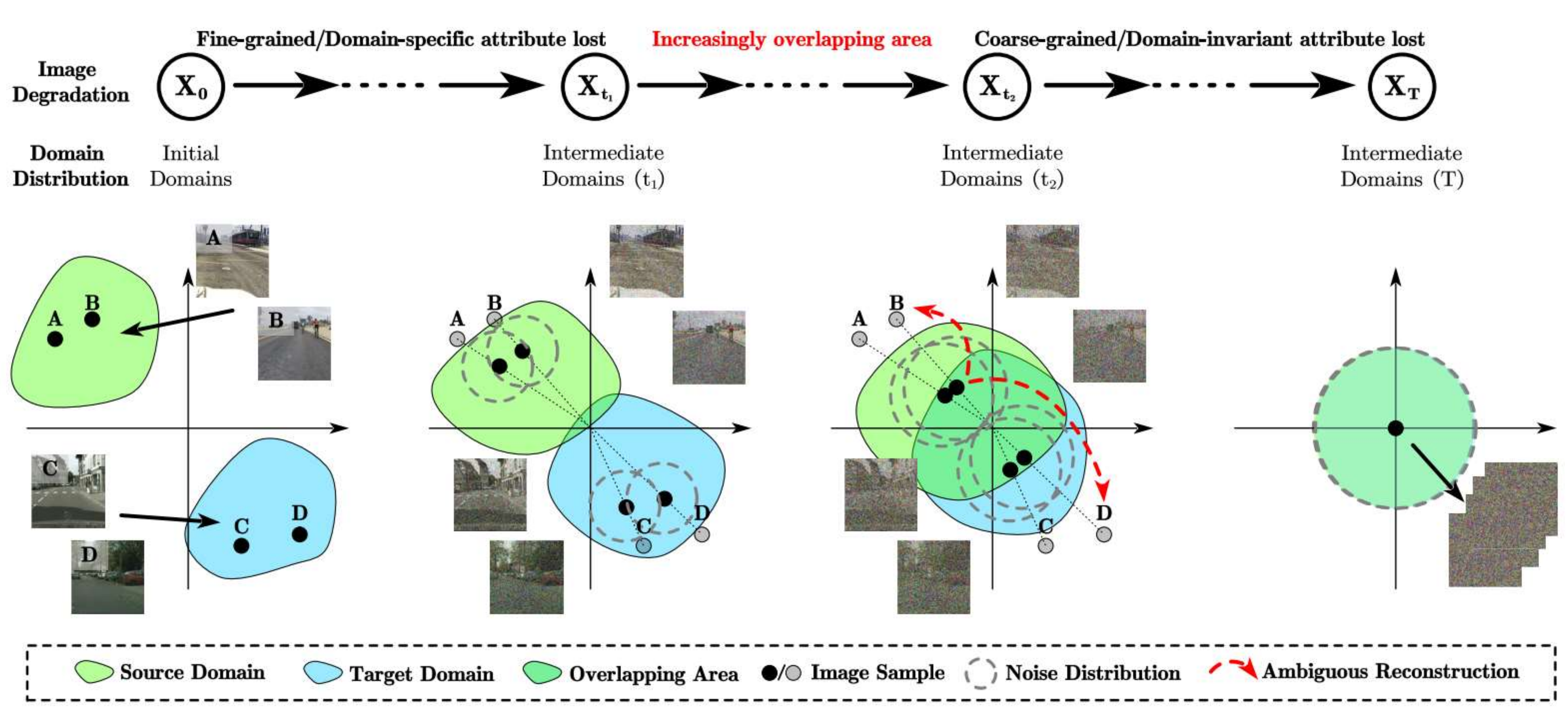
Introduction

Goal. Domain adaptive semantic segmentation aims to mitigate the performance degradation of segmentation networks by transferring knowledge from labeled source domains to unlabeled target domains.

Mainstream Paradigm. Self-training dominates UDA.

- Teacher–student pipeline: teacher (EMA of student) generates pseudo labels for unlabeled target images.
- Limitation: no explicit modeling of domain-shared feature extraction.

Motivation. In diffusion forward process, fine-grained, domain-specific attributes such as texture are lost with less noise added (i.e., early time-steps), while coarse-grained, domain-invariant attributes such as shape are lost by adding more noise (i.e., late time-steps).



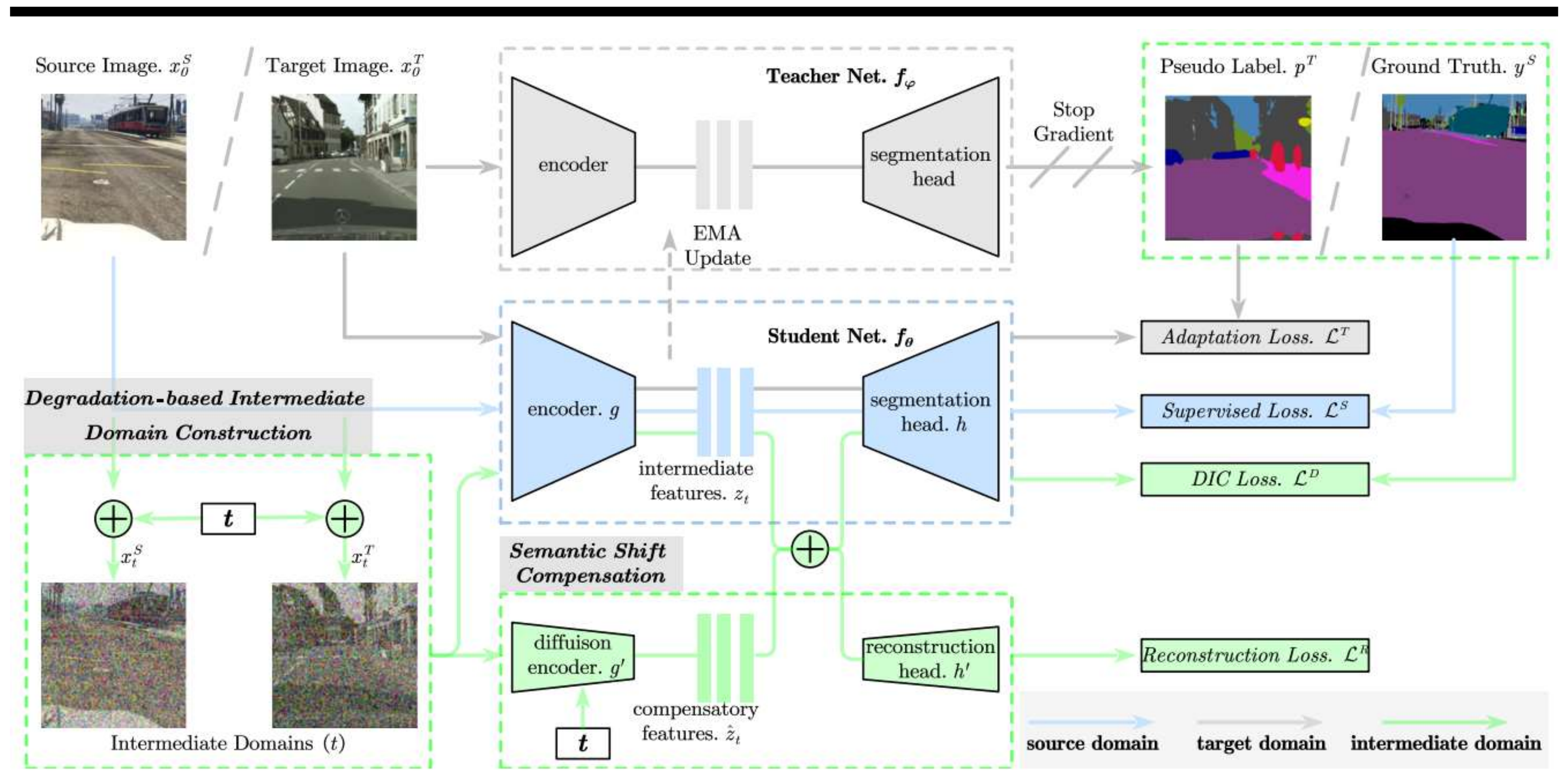
Inspiration. Simple image degradation operations can serve as effective priors for unsupervised domain bridging.

Theoretical Reinterpretation

Proposition (Attribute Loss and Time Step). *1) For each attribute Z_i, there exists a minimum time step $t(Z_i)$ such that Z_i is lost with degree τ at every $t \in \{t(Z_i), \dots, T\}$. 2) There exists a set $\{\beta_i\}_{i=1}^{T}$ such that $t(Z_i) > t(Z_j)$ whenever the distribution of $\|x_0 - g_i \cdot x_0\|$ first-order stochastically dominates that of $\|x_0 - g_j \cdot x_0\|$.*

- Fine-grained attributes vanish earlier than coarse ones, and lost attributes cannot be recovered at later steps.
- The growing overlap between degraded source/target distributions can be viewed as a domain-shared prior.

Method



Degradation-based Intermediate Domains: Formalize simple degradations as a forward diffusion process to create continuous intermediate domains:

$$X_T \rightarrow X_{T-1} \cdots X_t \xrightarrow{p_\theta(X_{t-1}|X_t)} X_{t-1} \cdots \rightarrow X_0 \sim \{\pi_s, \pi_t\}$$

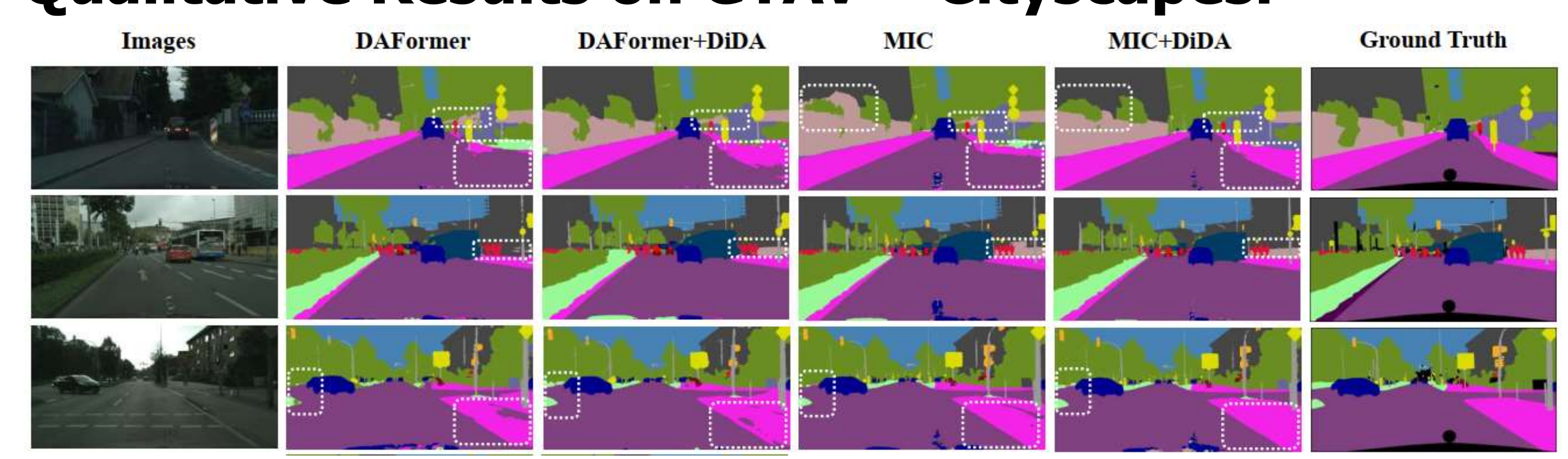
Semantic Shift Compensation: Add a time-conditioned diffusion encoder g' to disentangle t-specific shifts, supervised by reconstruction ($\mathcal{L}^R$) and Degraded Image Consistency ($\mathcal{L}^{DIC}$), with a reconstruction head h'.

Experiments

Quantitative Results on Different Benchmarks.

Method	GTA.→CS.		SYN.→CS.		CS.→ACDC
	C	T	C	T	T
DAFormer [38]	56.0	68.3	54.7	60.9	55.4
+DiDA	$58.3_{\uparrow 2.3}$	$70.3_{\uparrow 2.0}$	$57.6_{\uparrow 2.9}$	$63.1_{\uparrow 2.2}$	$59.1_{\uparrow 3.7}$
HRDA [39]	63.0	73.8	61.2	65.8	68.0
+DiDA	$64.3_{\uparrow 1.3}$	$75.4_{\uparrow 1.6}$	$62.6_{\uparrow 1.4}$	$67.8_{\uparrow 2.0}$	$70.7_{\uparrow 2.7}$
MIC [40]	64.2	75.5*	62.4*	67.3	69.8*
+DiDA	$65.0_{\uparrow 0.8}$	$\mathbf{76.8}_{\uparrow 1.3}$	$63.5_{\uparrow 1.1}$	$\mathbf{68.6}_{\uparrow 1.3}$	$\mathbf{72.1}_{\uparrow 2.3}$

Qualitative Results on GTAv→ Cityscapes.



Performance Variation with the Degraded Level.