Learning to See in Novel Viewpoints and Domains

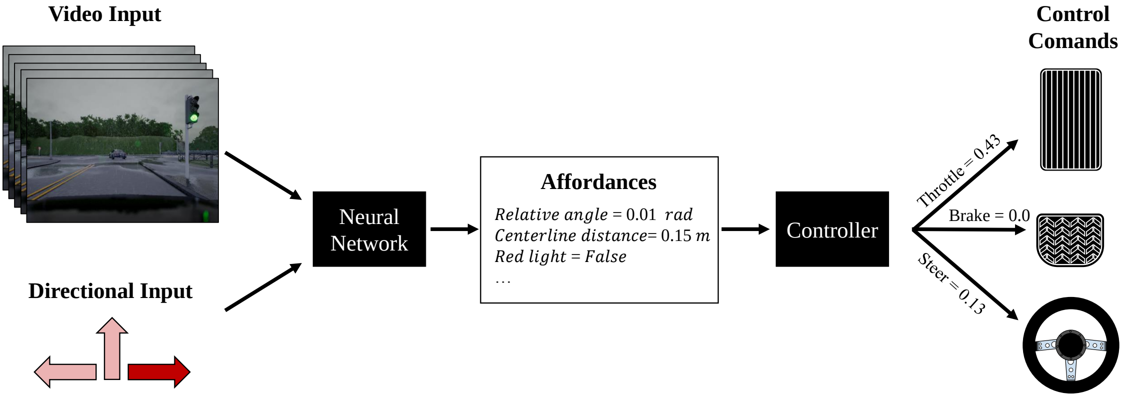
Andreas Geiger

Autonomous Vision Group
University of Tübingen / MPI for Intelligent Systems

June 16, 2018

University of Tübingen
MPI for Intelligent Systems
Autonomous Vision Group

Conditional Affordance Learning



Sauer, Savinov and Geiger: Conditional Affordance Learning for Driving in Urban Environments. CoRL, 2018.

NoVA:
Learning to See in Novel Viewpoints and Domains

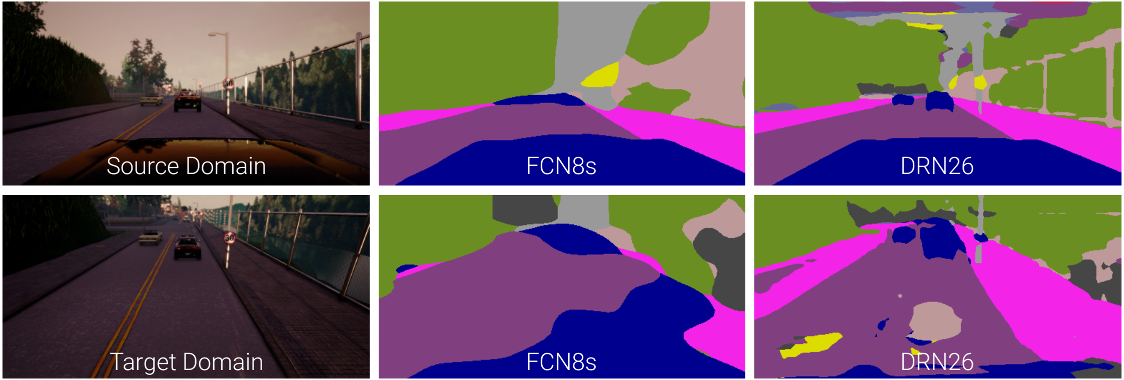
[Coors & Geiger, 2019]

Are all autonomous vehicles created equal?

Are all autonomous vehicles created equal?

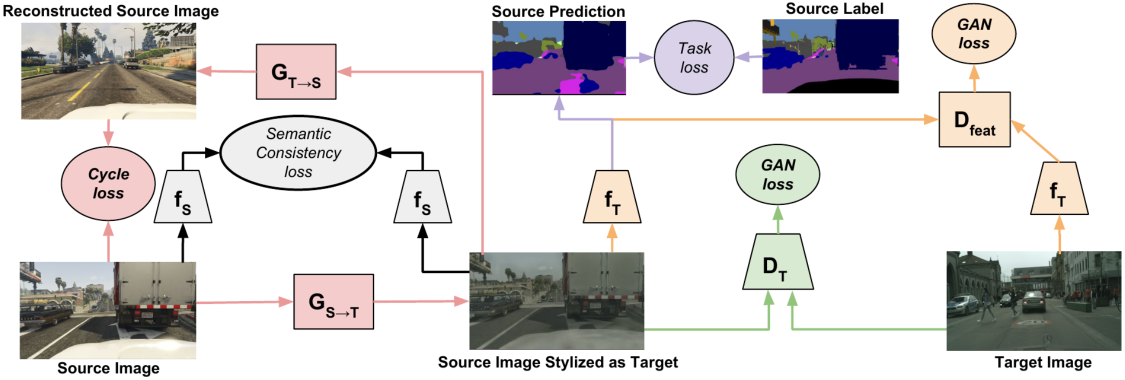
- Senor setups **vary significantly** depending on the **type** of autonomous vehicle

Generalization is Challenging



- Semantic Segmentation networks trained on source domain **do not generalize** well to novel viewpoints (target domain)

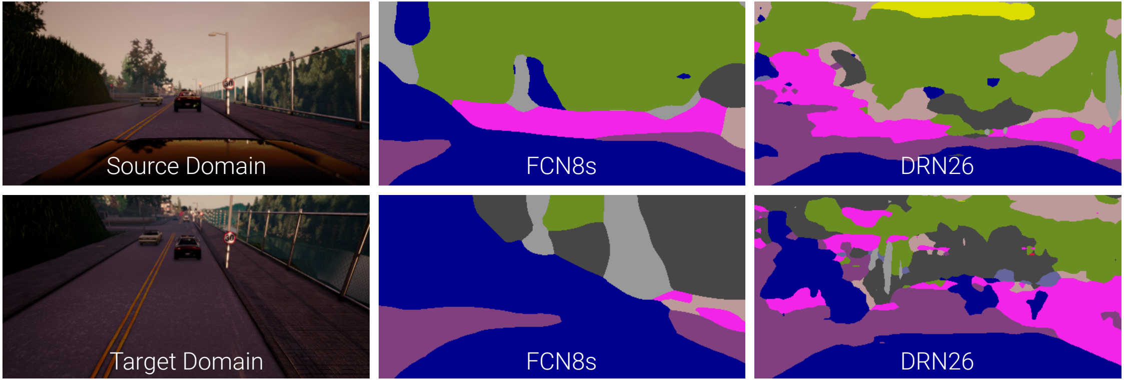
Cycle-Consistent Adversarial Domain Adaptation



- CyCADA addresses **domain adaptation** but not viewpoint adaptation

Hoffman, Tzeng, Park, Zhu, Isola, Saenko, Efros and Trevor Darrell: CyCADA: Cycle-Consistent Adversarial Domain Adaptation. ICML, 2018.

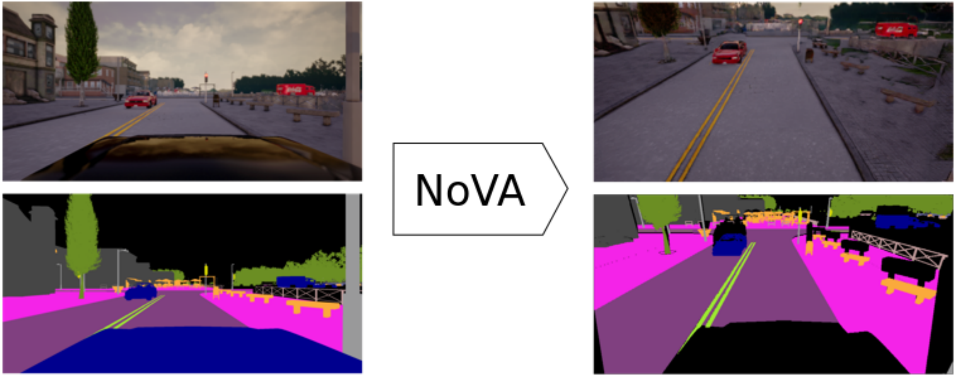
Cycle-Consistent Adversarial Domain Adaptation



- **CyCADA** does not handle viewpoint changes
- Leads to significant performance drop also for source domain

Hoffman, Tzeng, Park, Zhu, Isola, Saenko, Efros and Trevor Darrell: CyCADA: Cycle-Consistent Adversarial Domain Adaptation. ICML, 2018.

Novel Viewpoint and Domain Adapation



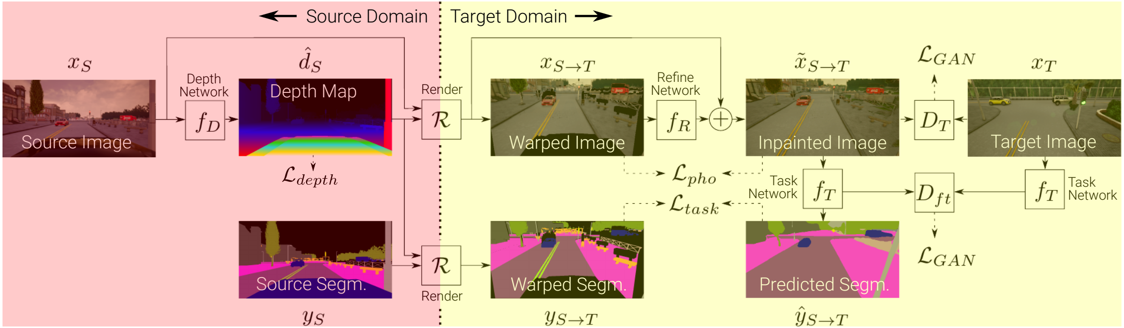
Source domain

Target domain (no labels)

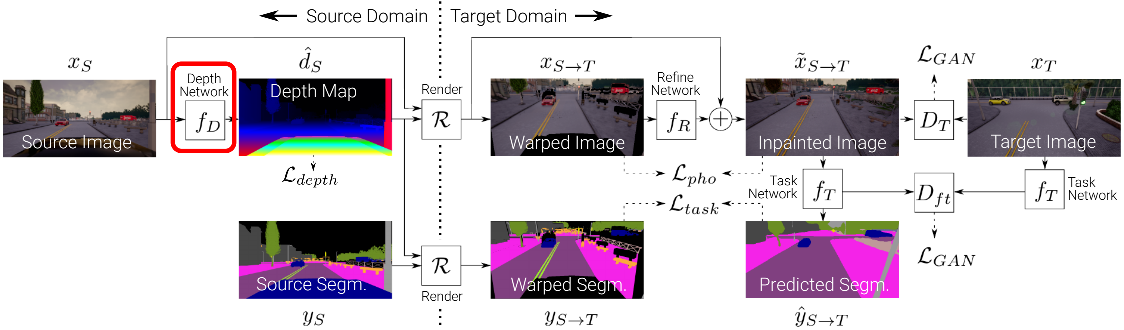
- **NoVA:** Geometry-aware image and semantic label translation
- **Key Idea:** Decompose problem (reconstruct, warp, inpaint, stylize)

Model Overview

Model Overview



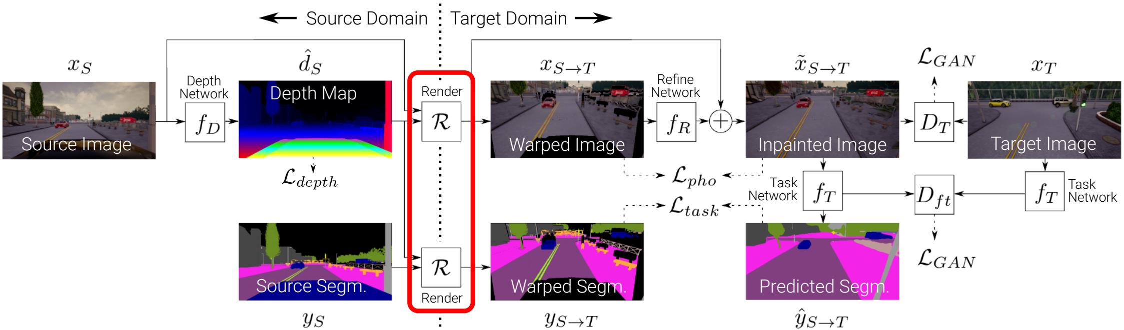
Depth Estimation



Predict depth from a single image:

- Can be trained either **supervised** using depth ground truth
- Or **self-supervised** using a second frame for supervision [Godard, CVPR 2017]

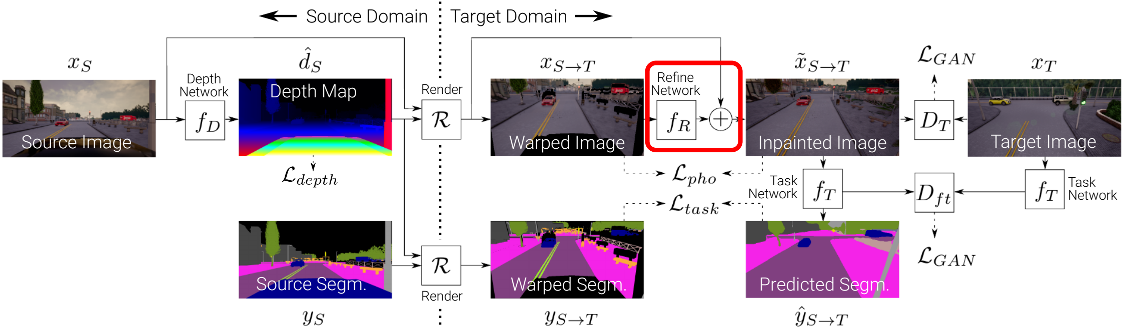
Rendering



Warp depth map and segmentation into target view:

- Forward **splatting** $\mathbf{K}_T\,\mathbf{T}\,\mathbf{K}_S^{-1}$
- Soft **z-buffering** (weight based on inverse depth)

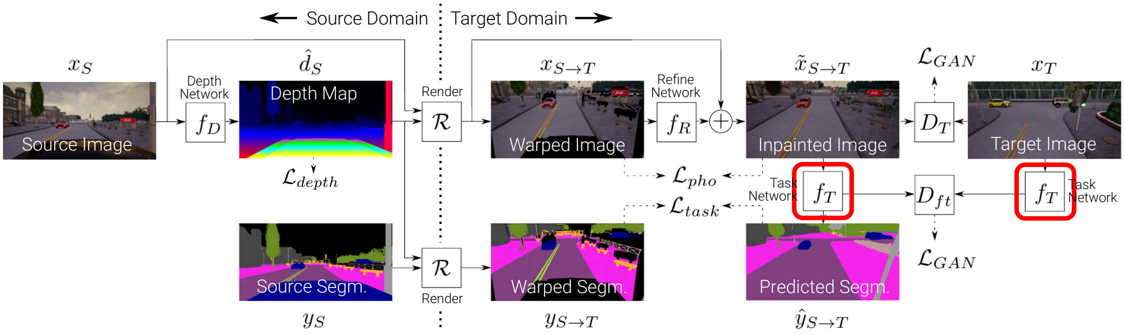
Inpainting and Refinement



Refine warped image:

- **Inpaint** occluded areas and **stylize** in style of target domain
- Residual network

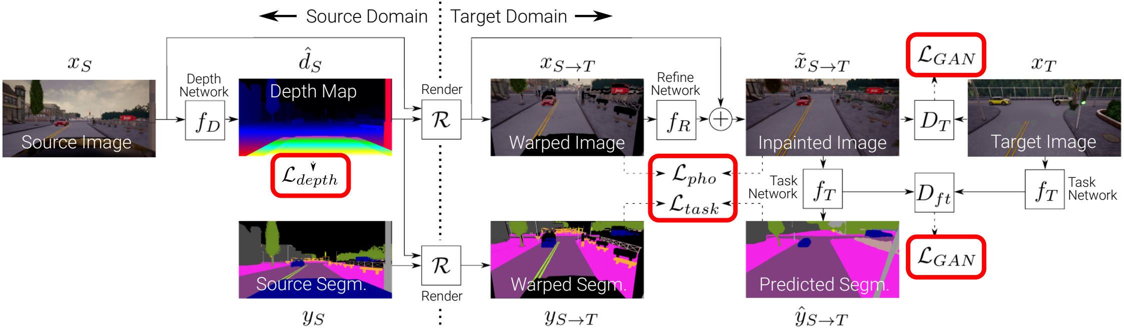
Task Network



Semantic segmentation task network:

- FCN8s and DRN26, **cross-entropy loss** applied to valid pixels
- **Adversarial losses** encourage similarity of image and semantic spaces

Loss Function



Loss function: $\mathcal{L}_{NoVA} = \mathcal{L}_{depth} + \mathcal{L}_{pho} + \mathcal{L}_{task} + \mathcal{L}_{GAN}$

Experimental Evaluation

Viewpoint Adaptation

CARLA Car $\rightarrow$ CARLA Truck

Viewpoint Adaptation

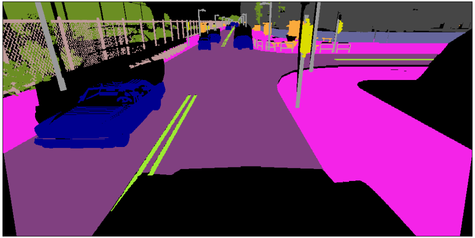
Method	mIoU
Source Only	26.54
SceneAdapt [Mauro et al., AVSS 2018]	26.63
CyCADA [Hoffman et al., ICML 2018]	10.57
CyCADA [Hoffman et al., ICML 2018] + *trgt-labels*	16.31
SPLAT [Tzeng et al., 2018]	13.63
SPLAT [Tzeng et al., 2018] + *trgt-labels*	18.81
NoVA$_{mono-self}$	42.54
NoVA$_{mono-sup}$	45.27
NoVA$_{stereo-sup}$	49.67
NoVA$_{GT}$	**51.89**
Target Oracle	52.72

Viewpoint Adaptation

Source Image

Warped Image

Refined Image



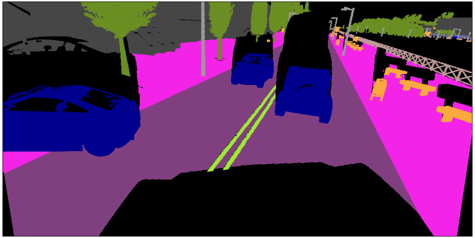
Warped Labels

Viewpoint Adaptation

Source Image

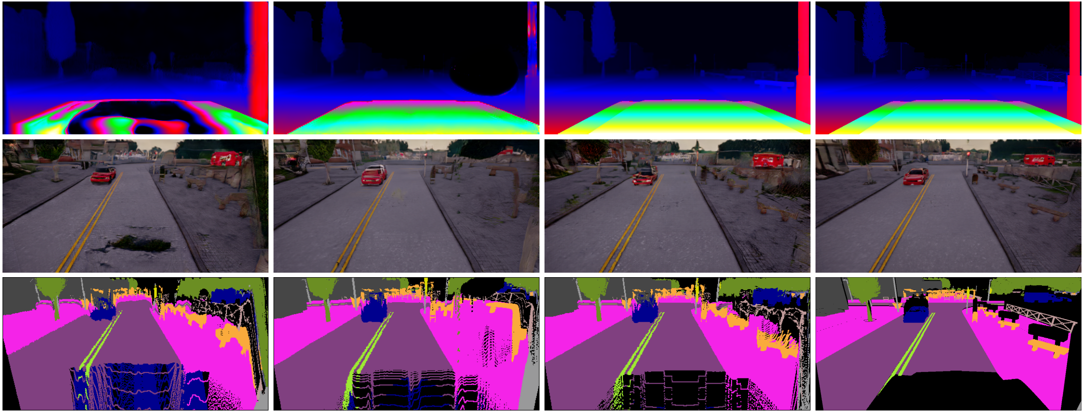
Warped Image

Refined Image



Warped Labels

Viewpoint Adaptation



NoVA$_{mono-self}$ NoVA$_{mono-sup}$ NoVA$_{stereo-sup}$ NoVA$_{GT}$

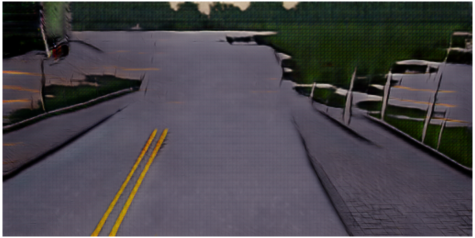
NoVA Performance for Different Depth Estimators:

- NoVA works with self-supervised or supervised depth predictions

Viewpoint Adaptation

Source Image

NoVA$_{mono-self}$

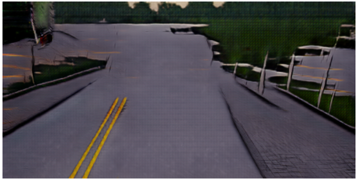


CyCADA

SPLAT

Viewpoint Adaptation

Source



Translated

Reconstructed

CyCADA on Sim2Sim:

- CyCADA translates incorrectly but reconstructs correctly
- Learns to encode source image semantics in the noise of translated image

Viewpoint and Domain Adaptation

CARLA Truck $\rightarrow$ Cityscapes Car

Viewpoint and Domain Adaptation

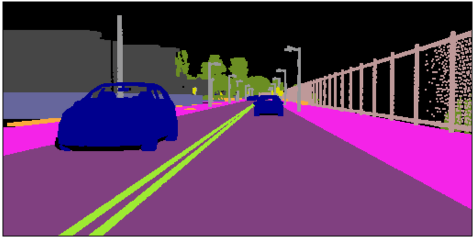
Method	mIoU
Source Only	18.84
SceneAdapt [Mauro et al., AVSS 2018]	11.54
CyCADA [Hoffman et al., ICML 2018]	19.26
SPLAT [Tzeng et al., 2018]	21.01
NoVA$_{mono-self}$	30.23
NoVA$_{mono-sup}$	34.36
NoVA$_{stereo-sup}$	32.96
NoVA$_{GT}$	**35.91**
Target Oracle	51.30

Viewpoint and Domain Adaptation

Source Image

Warped Image

Refined Image



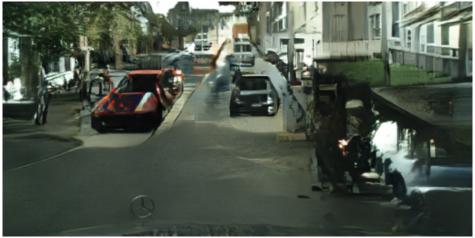
Warped Labels

Viewpoint and Domain Adaptation

Source Image

NoVA$_{mono-self}$

CyCADA



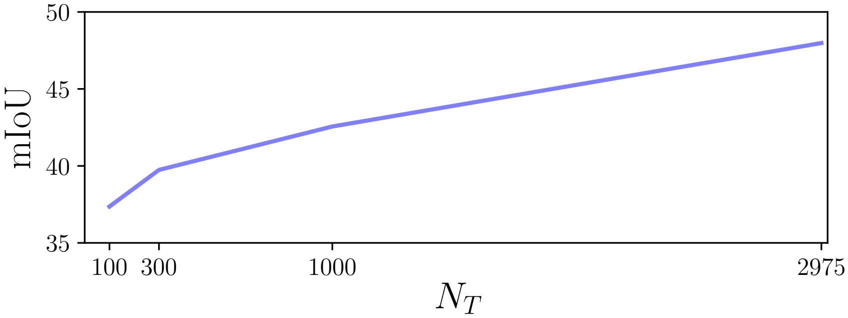
SPLAT

Ablation Study

Sim2Sim	mIoU
Source Only	26.54
+ Forward Warping	47.81
+ Residual Refinement	**51.89**

Sim2Real	mIoU
Source Only	18.84
+ Forward Warping	26.95
+ Residual Refinement	**35.91**

Semi-Supervised Viewpoint and Domain Adaptation



Semi-Supervised Adaptation on Sim2Real:

- Combination of $N_S = 30,000$ translated images with N_T labeled target images

Video

Summary

This work:

- Semantic segmentation networks **do not generalize** to novel viewpoints
- Existing domain adaptation techniques cannot address the problem
- NoVA is a new model for **viewpoint and domain adaptation**
  - **Geometry-aware** image and label translation
  - **Decompose** problem: depth estimation, warping, inpainting, domain

Future work:

- Warpings and inpaintings sometimes still exhibit artifacts
- Current results are not temporally consistent

Thank you!

http://autonomousvision.github.io

European Research Council
Established by the European Commission

Federal Ministry
of Education
and Research

Microsoft Research