DEN: Disentangling and Exchanging Network for Depth Completion

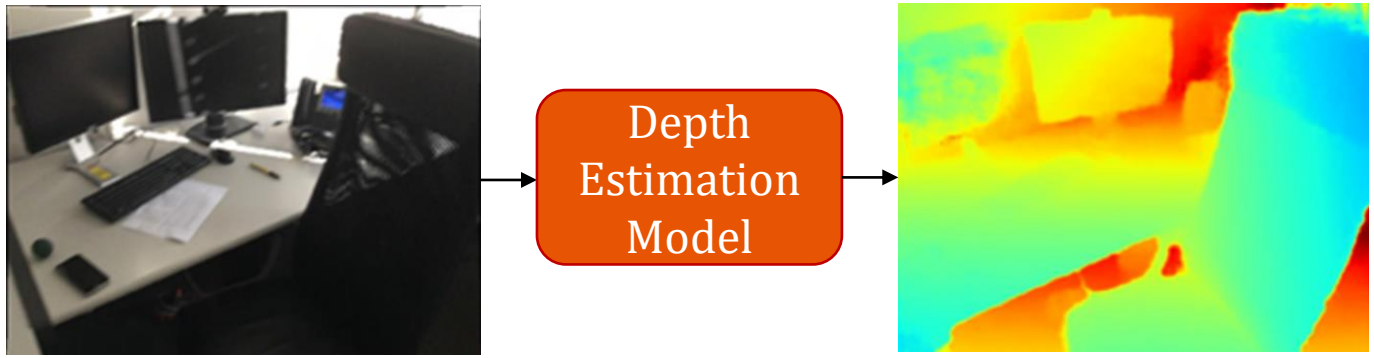
You-Feng Wu*, Vu-Hoang Tran†, Ting-Wei Chang‡, Wei-Chen Chiu‡, and Ching-Chun Huang‡

- **Intro**
  - Challenges of Depth Estimation
  - Our Setting
  - Overview
- **Previous Works**
  - Depth Representation Related
  - Disentangling Network Related
- **Method**
  - Depth Representation
  - Network Architecture
  - Criterion Design
- **Experiment**

Intro

Challenges of Depth Estimation

Monocular Depth Estimation

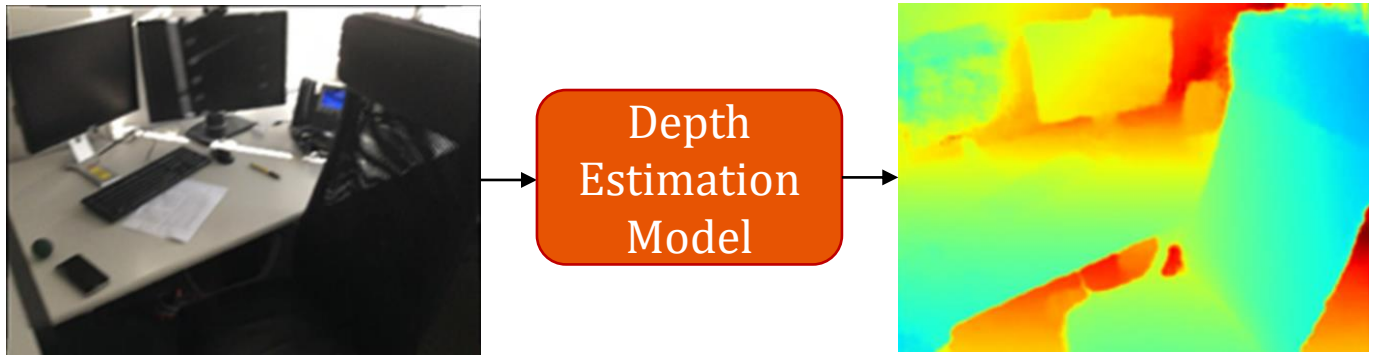


Challenges

- ☹ Spatial Scale Offset
- ☹ RGB image texture influence
- ☹ Mixed Depth Pixel

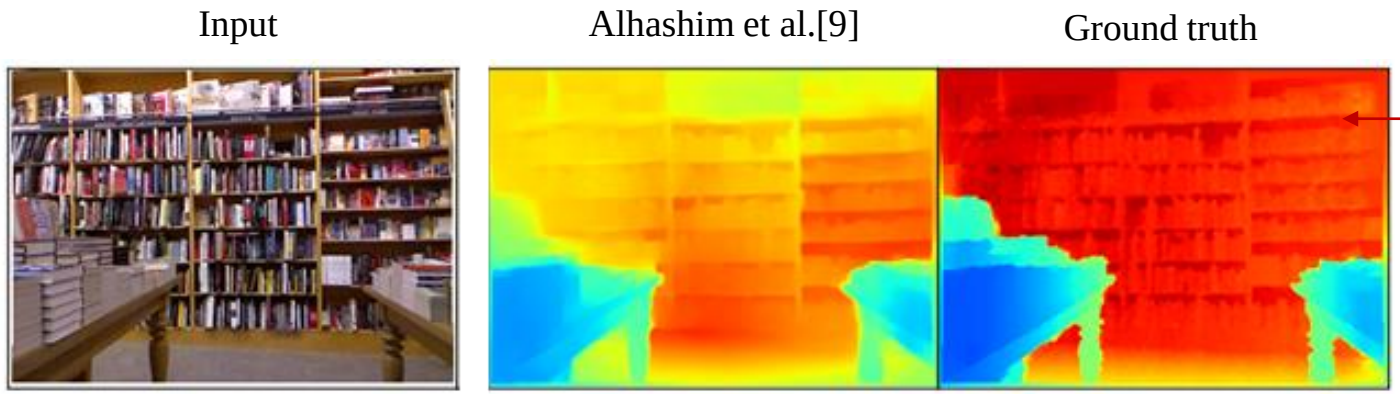
Challenges of Depth Estimation

Monocular Depth Estimation



Challenges

- ☹ **Spatial Scale Offset**
- ☹ RGB image texture influence
- ☹ Mixed Depth Pixel

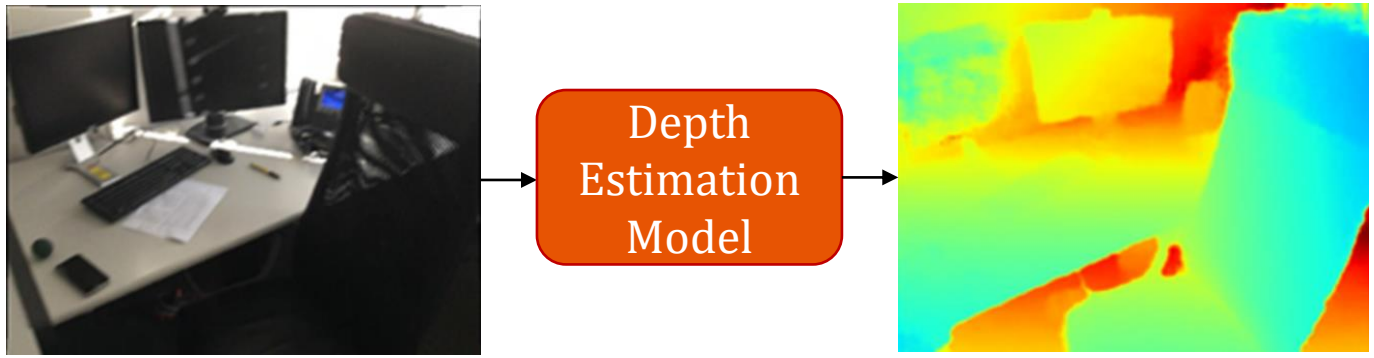


Model predicts incorrectly due to scale difference compare to ground truth image.

[9] High Quality Monocular Depth Estimation via Transfer Learning, arXiv, 2019

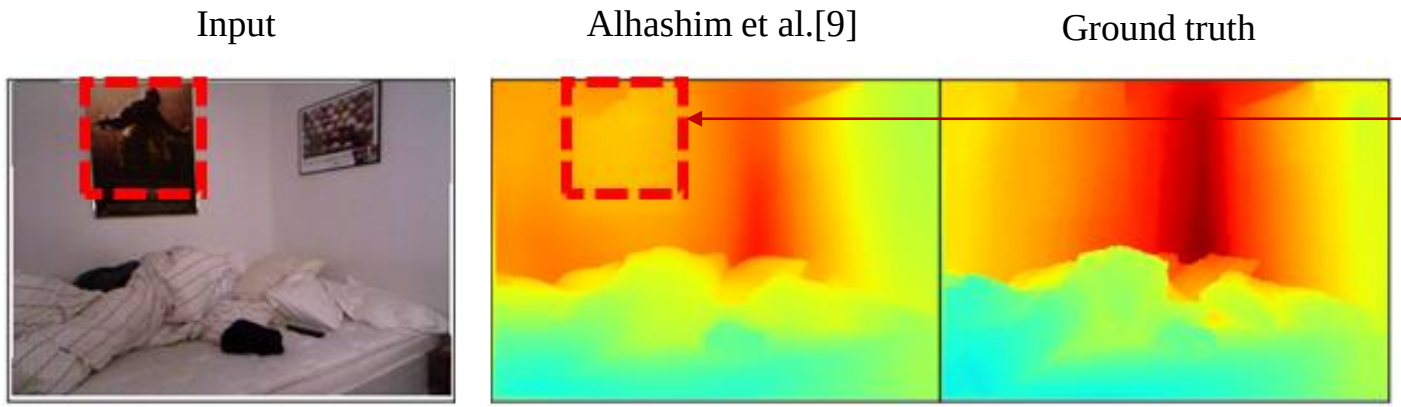
Challenges of Depth Estimation

Monocular Depth Estimation



Challenges

- ☹ Spatial Scale Offset
- ☹ **RGB image texture influence**
- ☹ Mixed Depth Pixel

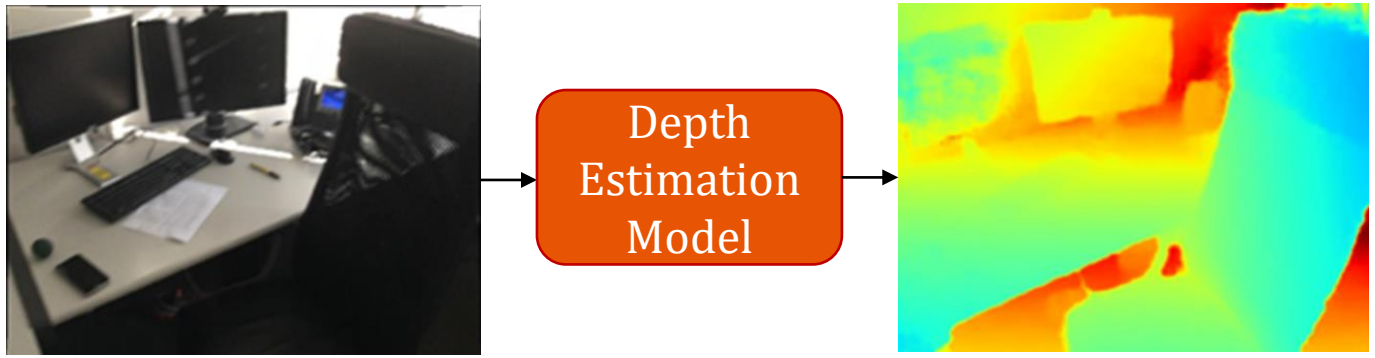


Model is affected by the painting on the wall and predict undesired result.

[9] High Quality Monocular Depth Estimation via Transfer Learning, arXiv, 2019

Challenges of Depth Estimation

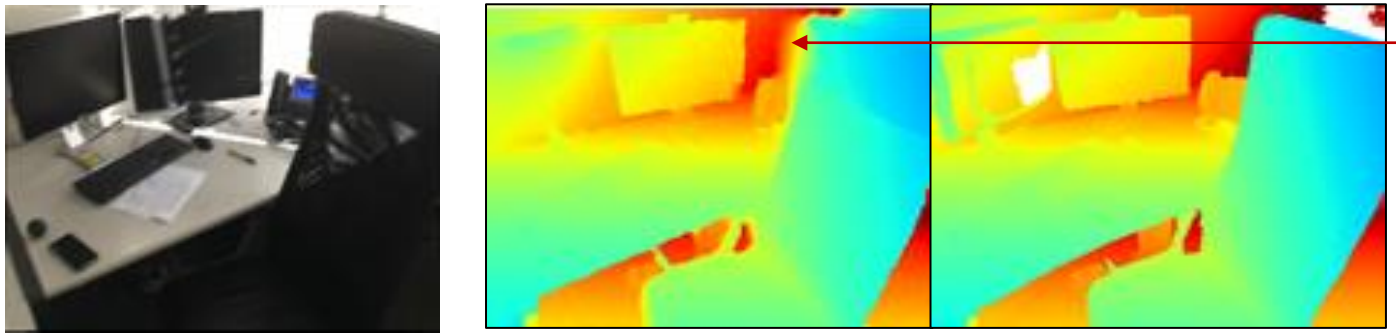
Monocular Depth Estimation



Challenges

- ☹ Spatial Scale Offset
- ☹ RGB image texture influence
- ☹ **Mixed Depth Pixel**

Input | Zhang et al.[13] | Ground truth

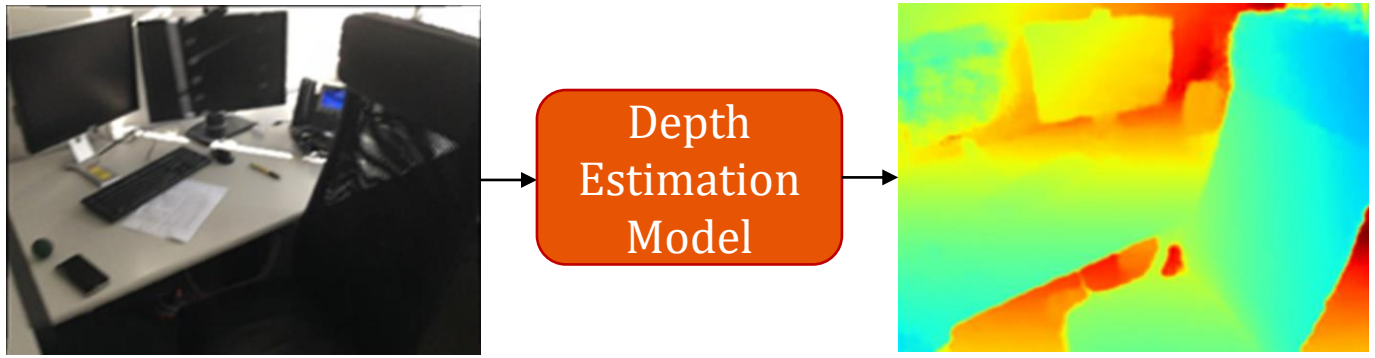


Undesired prediction between foreground and background

[13] Deep Depth Completion of a Single RGB-D Image, CVPR, 2018

Challenges of Depth Estimation

Monocular Depth Estimation

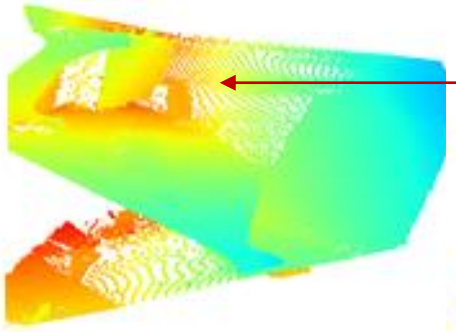


Challenges

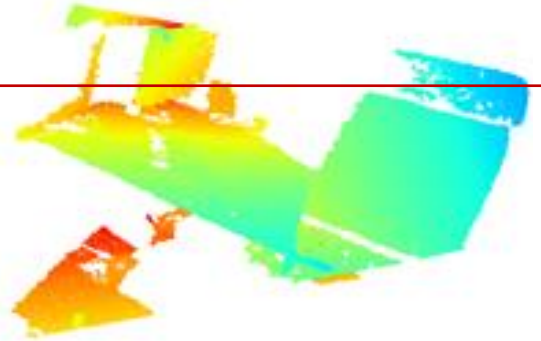
- ☹ Spatial Scale Offset
- ☹ RGB image texture influence
- ☹ **Mixed Depth Pixel**

Input

Zhang et al.[13]



Ground truth

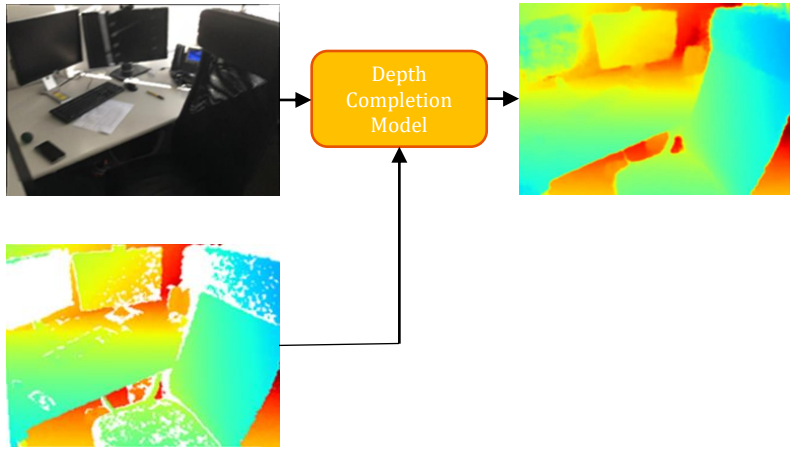


We can observe this problem clearer in 3D projection

[13] Deep Depth Completion of a Single RGB-D Image, CVPR, 2018

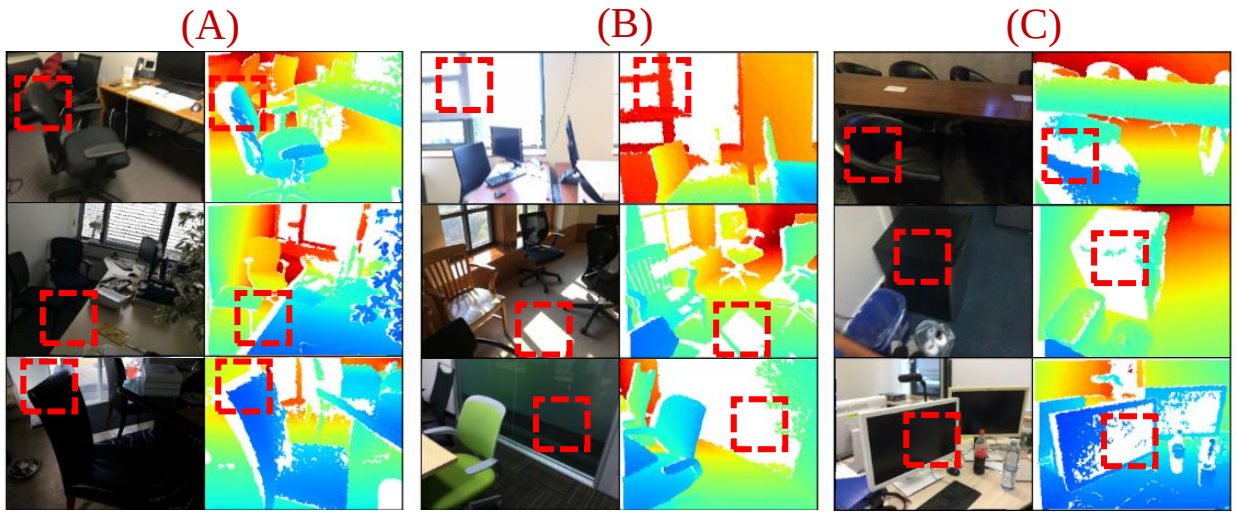
Our Setting

Depth Completion



Fail Cases of Commercial Depth Camera

- Significant depth difference between foreground and background (A) [24]
- Shiny, bright, transparent (B), and distant surfaces (C) [25]



Motivation

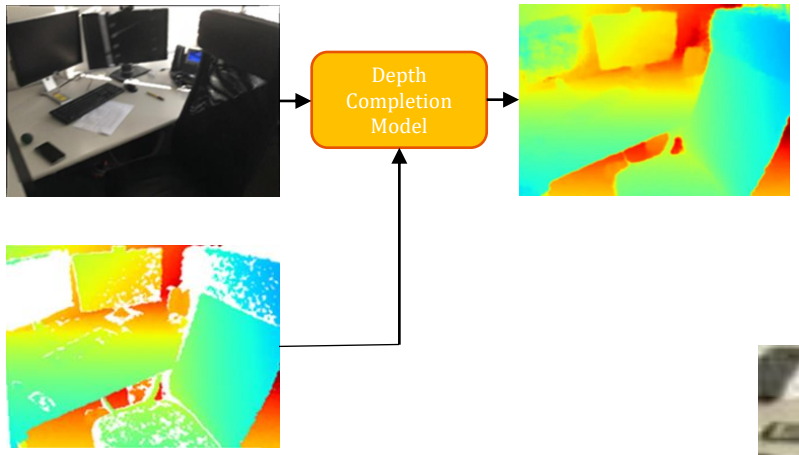
We aims to design an algorithm that can leverage on knowledge of incomplete depth map generated by commercial depth camera for

- More Accurate results compare to monocular depth estimation
- More Practical than design or purchase a higher-level depth camera for better quality depth map

[24] Superpixel-based depth map inpainting for RGB-D view synthesis, ICIP, 2015
[25] Reconstructing Scenes with Mirror and Glass Surfaces, TOG, 2018

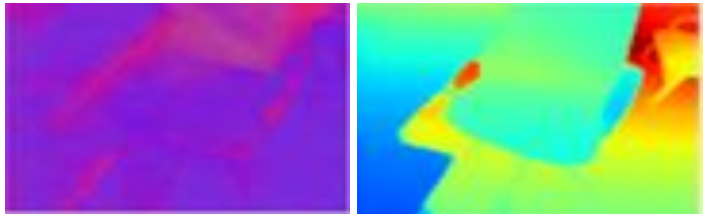
Our Setting

Depth Completion



Input: RGB-D

Annotations

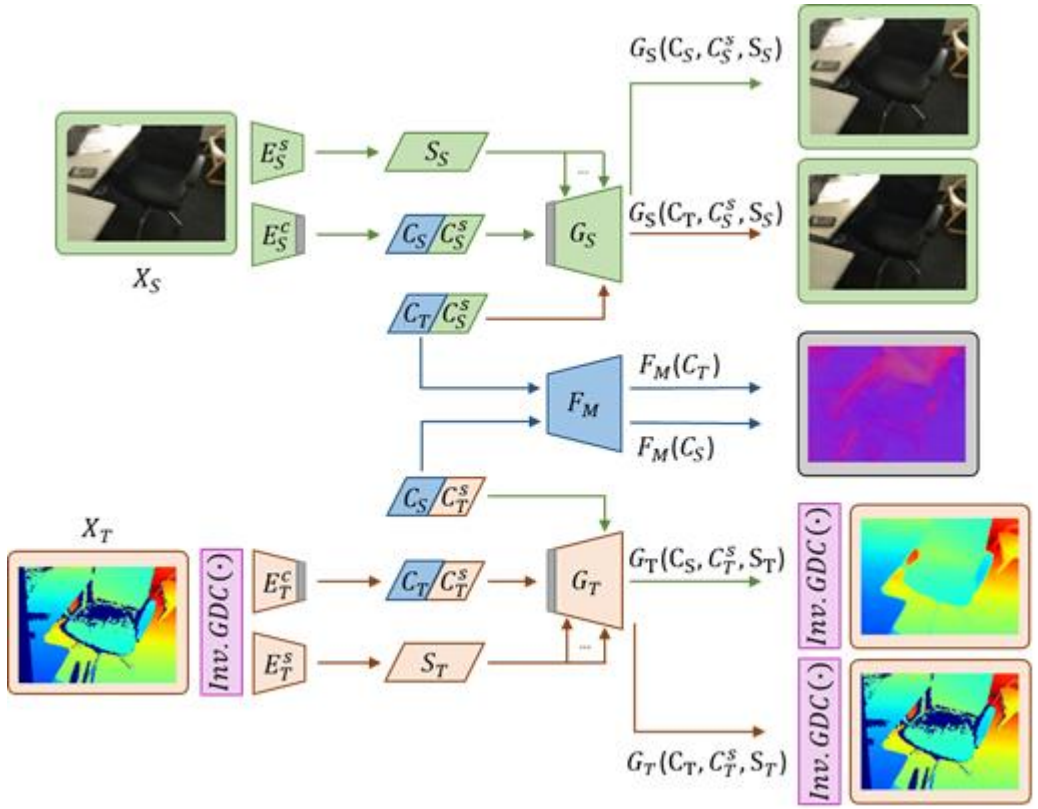


Dataset we used

- **ScanNet [7]**
  - Used on several 3D scene understanding task
  - Provide 1. **color image**, 2. **incomplete depth image**, and annotated with 3. **ground truth depth** and 4. **surface normal**, etc.
  - Train: 59743 pairs of data from 1000 scenes
  - Test: another 500 pairs from other scenes

[7]: ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes

Overview



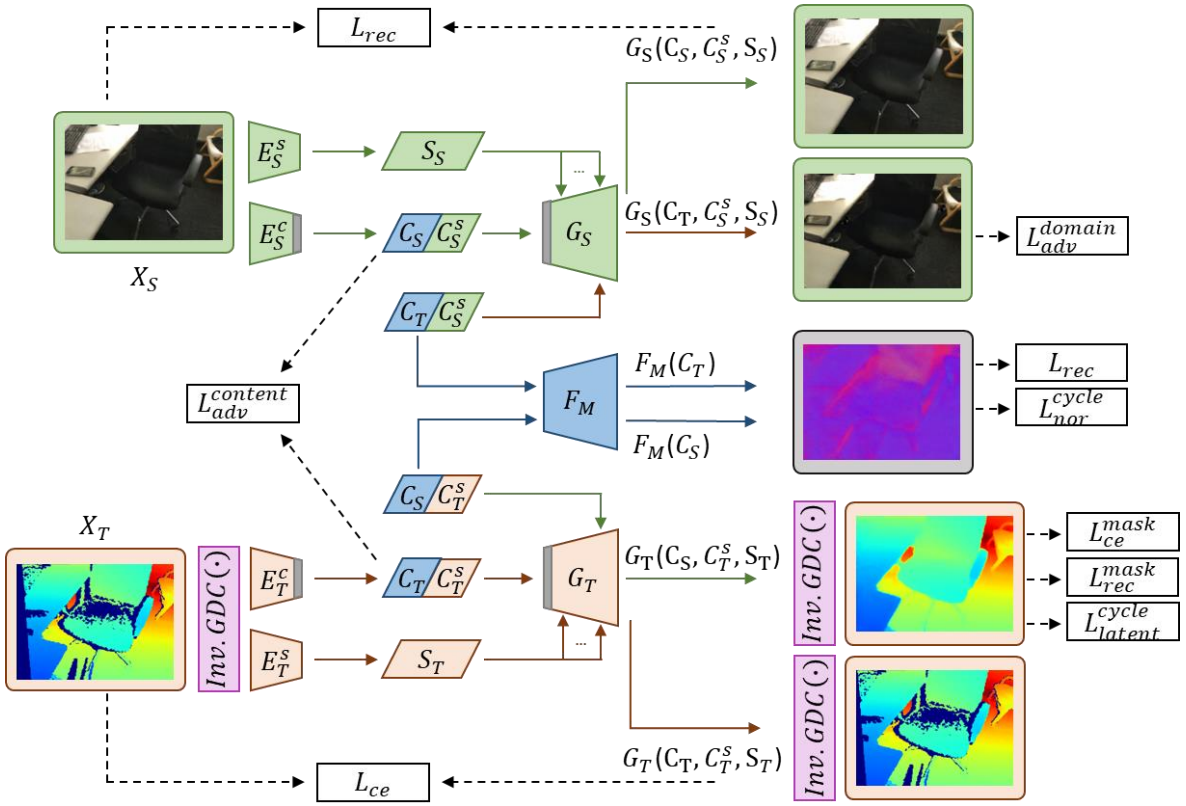
Model Architecture

- Disentangled Representation Learning
- Domain Adaptation
- Feature exchange across domains

Depth Representation

- General Depth Representation

Overview



Model Architecture

- **Disentangled Representation Learning**
- **Domain Adaptation**
- **Feature exchange across domains**

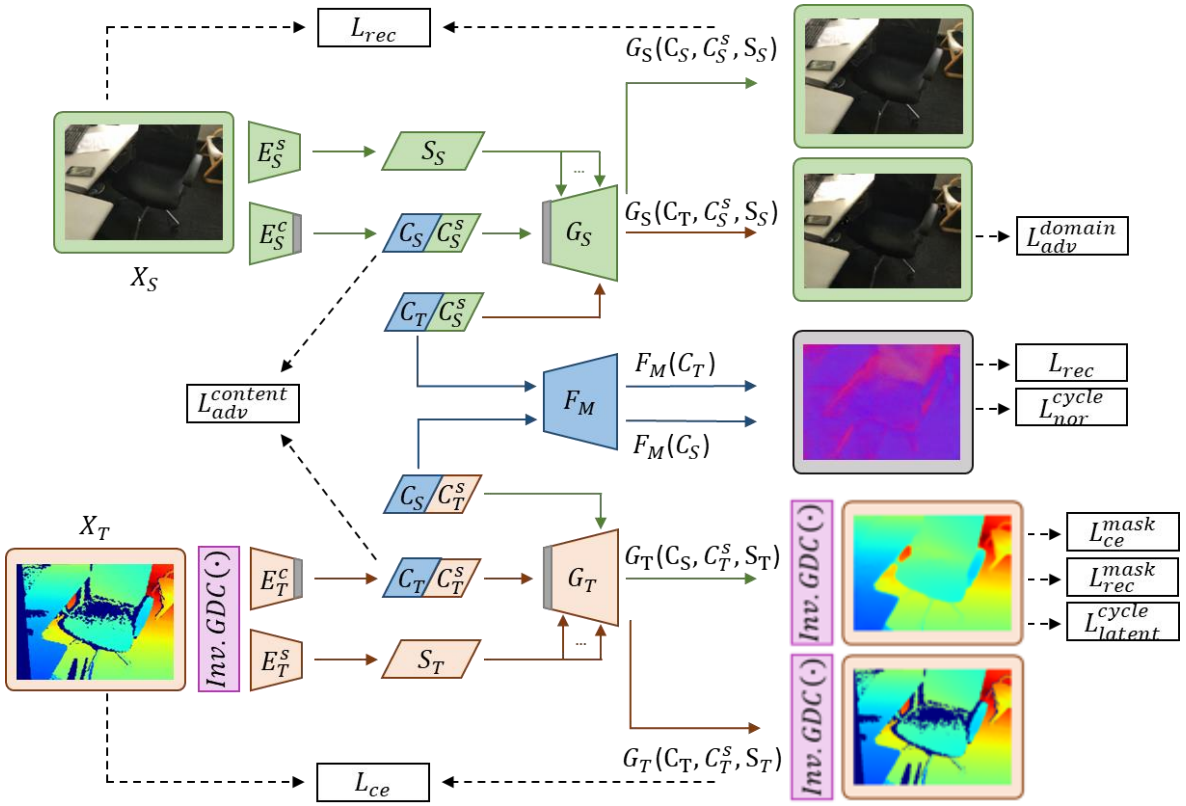
Depth Representation

- General Depth Representation

Challenges

- ☹ ~~Spatial Scale Offset~~
- ☹ ~~RGB image texture influence~~
- ☹ Mixed Depth Pixel

Overview



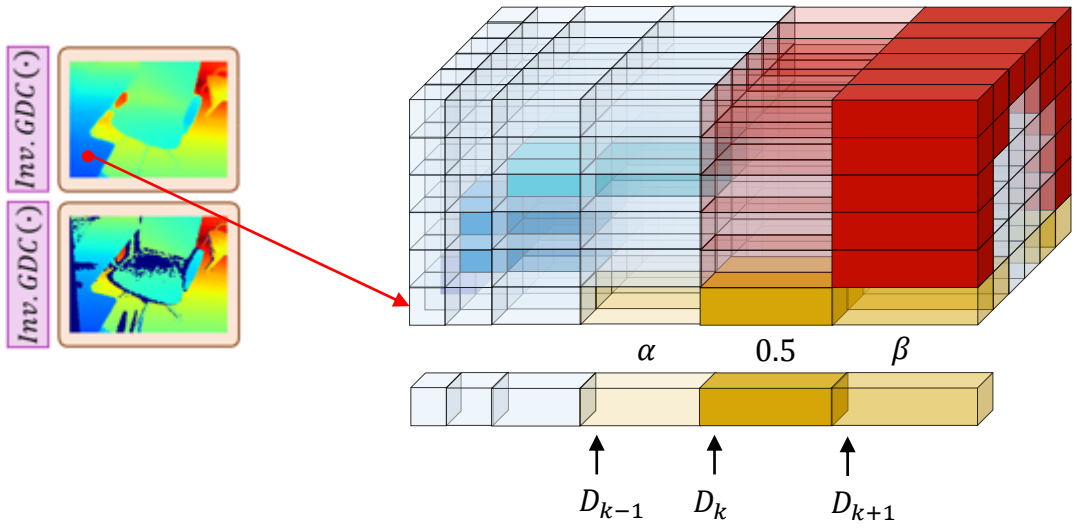
Model Architecture

- Disentangled Representation Learning
- Domain Adaptation
- Feature exchange across domains

Depth Representation

- **General Depth Coefficient**

Overview



$$d=\alpha * D_{k-1} + 0.5 * D_k + \beta * D_{k+1}$$

Model Architecture

- Disentangled Representation Learning
- Domain Adaptation
- Feature exchange across domains

Depth Representation

- **General Depth Coefficient**

Challenges

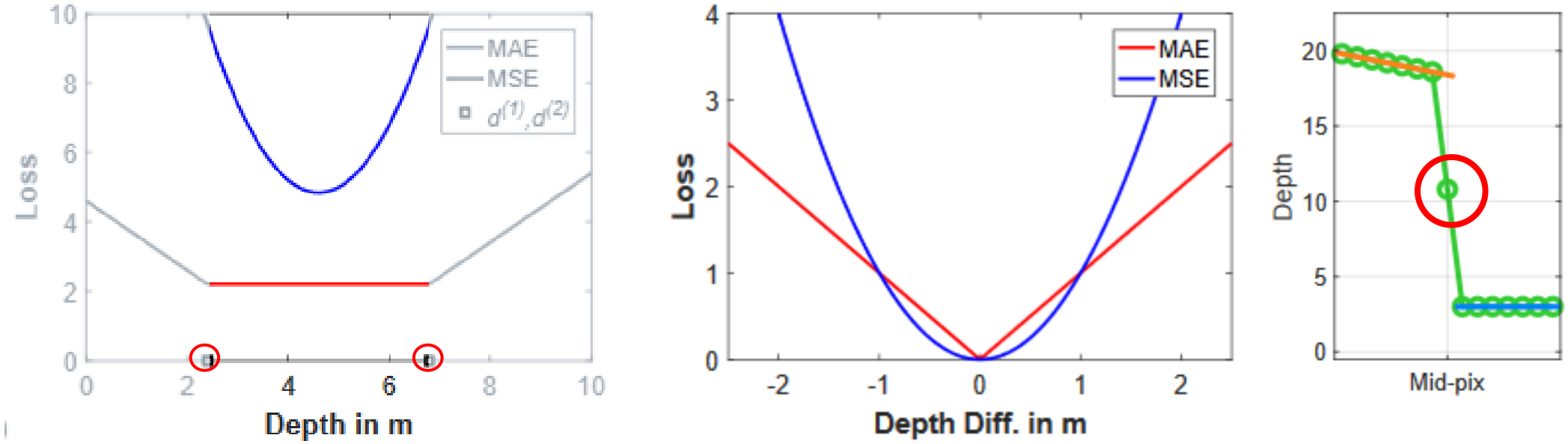
- ☹ ~~Spatial Scale Offset~~
- ☹ ~~RGB image texture influence~~
- ☹ ~~Mixed Depth Pixel~~

Previous Works

Depth Representation Related

Mixed Depth Pixel

- Many methods model depth estimation as a **regression problem** [12,18,20,35,44], which the model will prefer to generate mixed depth pixel for optimization.



[12] Deep depth from defocus: how can defocus blur improve 3d estimation using dense neural networks?, ECCV, 2018
[18] Depth Completion with Deep Geometry and Context Guidance, ICRA, 2019
[20] Reconstruction-based Pairwise Depth Dataset for Depth Image Enhancement Using CNN, ECCV, 2018
[35] Dfusenet: Deep fusion of rgb and sparse depth information for image guided dense depth completion, arXiv, 2019
[44] Parse geometry from a line: Monocular depth estimation with partial laser observation, ICRA, 2017

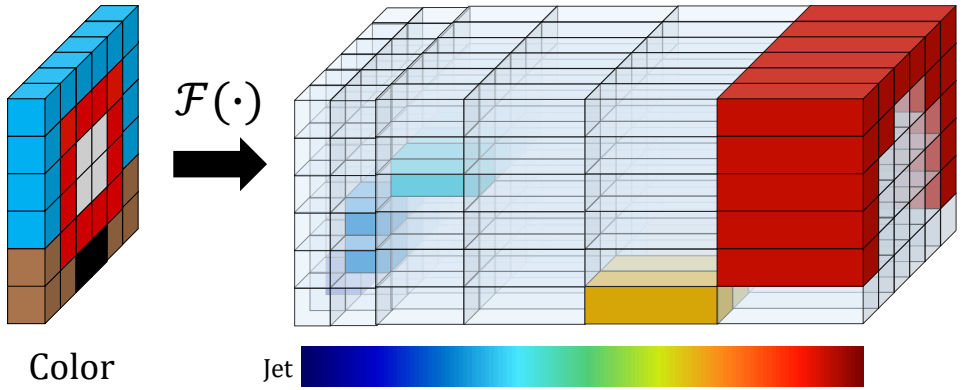
Depth Representation Related

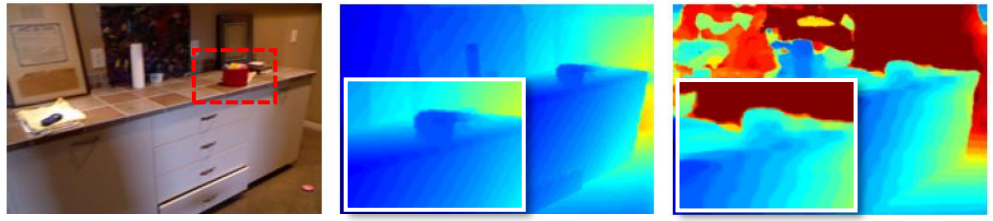
DORN [4]

- Depth Estimation: Regression Problem→Bin Classification Problem
- Loss: MSE/MAE →Cross Entropy

☹ Quantization Error

☹ Trade off between memory and precision





Input Ground truth DORN[4]

[4] Deep Ordinal Regression Network for Monocular Depth Estimation, CVPR, 2018

Depth Representation Related

DORN [4]

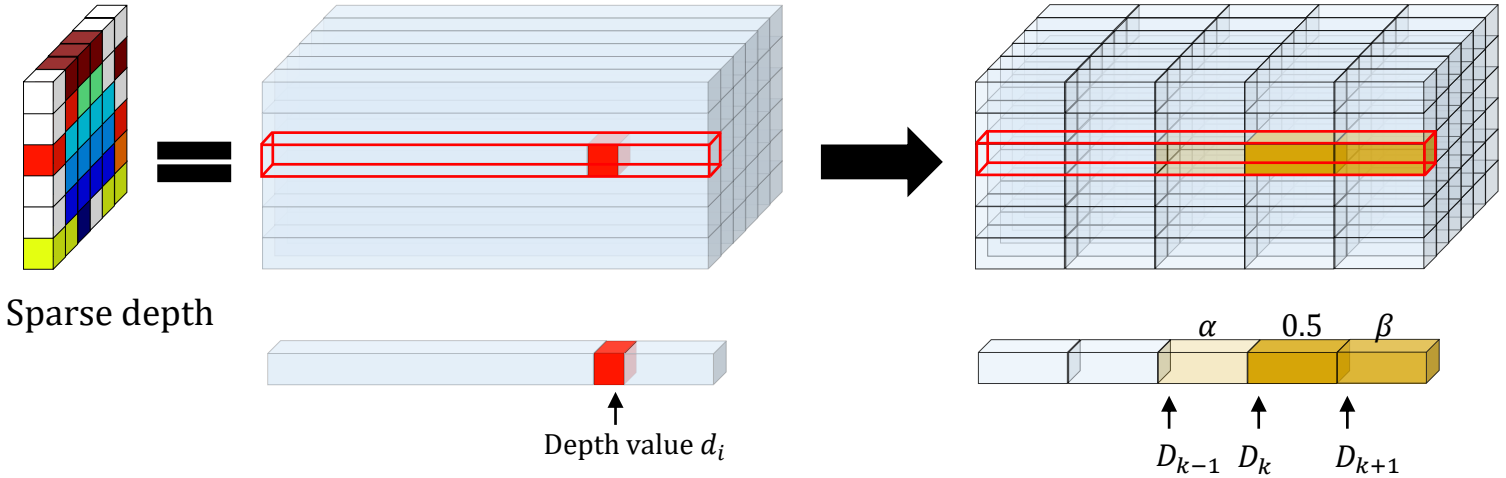
- Depth Estimation: Regression Problem→Bin Classification Problem
- Loss: MSE/MAE →Cross Entropy

☹ Quantization Error

☹ Trade off between memory and precision

Imran et al. [22]

- Proposed **<u>Depth Coefficient</u>** representation
- $d = \alpha * D_{k-1} + 0.5 * D_k + \beta * D_{k+1}$

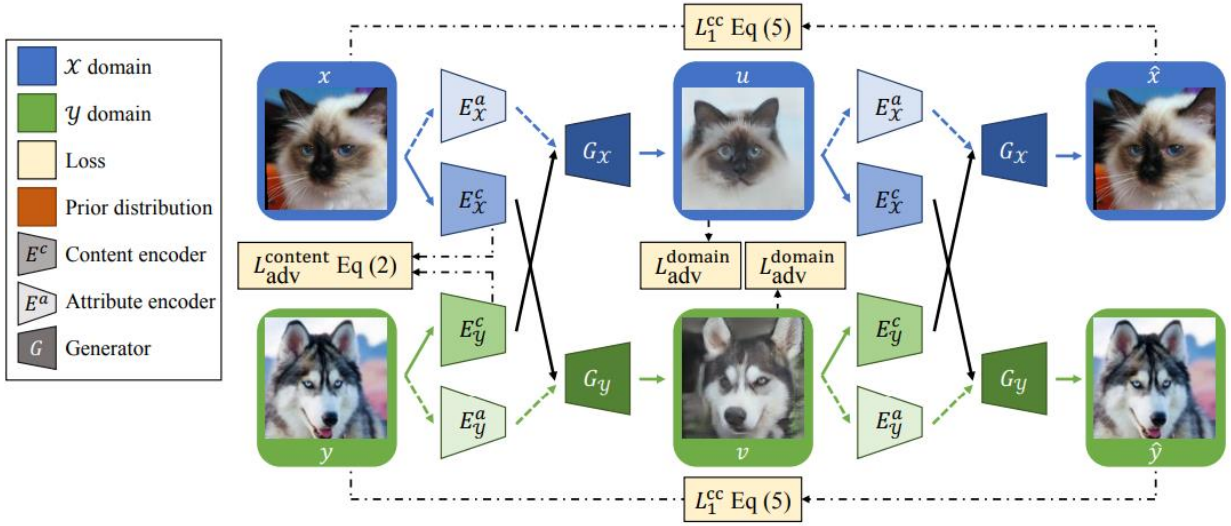


[4] Deep Ordinal Regression Network for Monocular Depth Estimation, CVPR, 2018
[22]Depth Coefficients for Depth Completion, arXiv, 2019

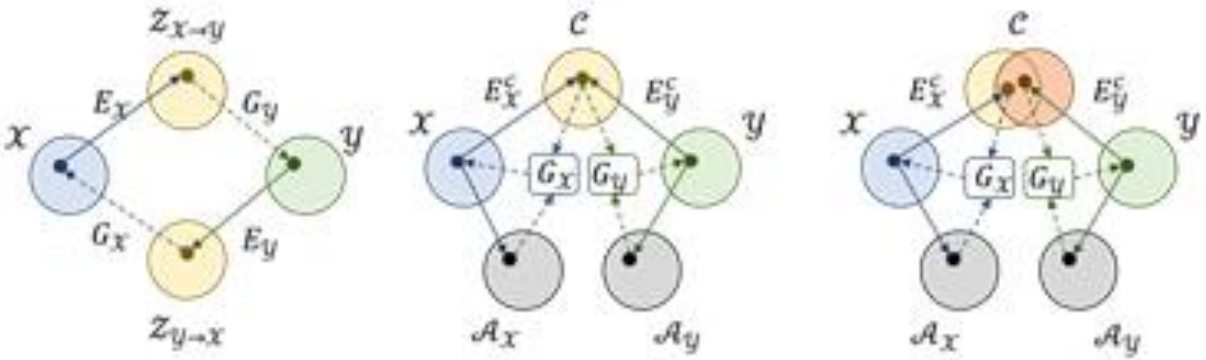
Disentangling Network Related

DRIT [41]

- Disentangling feature into content and attribute domain
- Aligning only content domain
- Exchanging attribute domain for style transferring



Comparison



CycleGAN [38] DRIT [41] Our

[38] Unpaired image-to-image translation using cycle-consistent adversarial networks, ICCV 2017
[41] Diverse image-to-image translation via disentangled representations, ECCV 2018

Method

Depth Representation

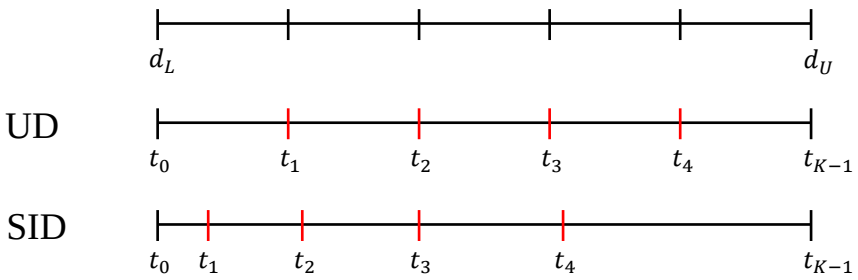
General Depth Coefficient

- Re-formula depth coefficients[22] based on spacing-increasing discretization (SID) [4]
  - Uniform Discretization (UD) [22]

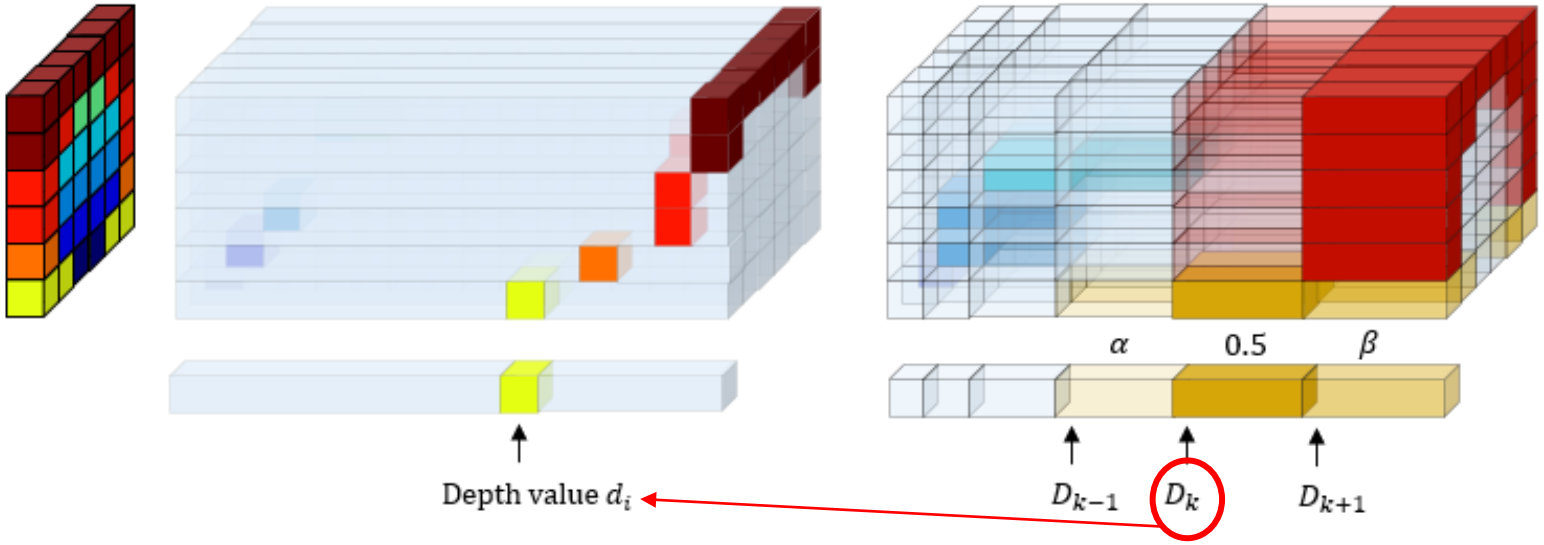
$$UD:\ t_i = d_L + (d_U - d_L) * i/K$$

  - Spacing Increasing Discretization (SID) [4]

$$SID:\ t_i = e^{\log(d_L) + \frac{log(d_U - d_L) * i}{K}}$$



Depth Value → Depth Coefficient



Nearest to d_i

$$c_i = \{0, \dots, 0, \alpha, 0.5, \beta, 0, \dots, 0\}$$

where $\alpha = \frac{d_i - 0.5(D_k + D_{k+1})}{D_{k-1} - D_{k+1}}$

$$\beta = 0.5 - \alpha$$

[4] Deep Ordinal Regression Network for Monocular Depth Estimation, CVPR, 2018
[22]Depth Coefficients for Depth Completion, arXiv, 2019

Depth Representation

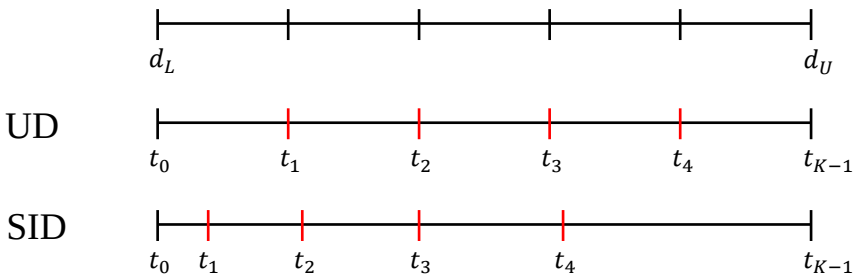
General Depth Coefficient

- Re-formula depth coefficients[22] based on spacing-increasing discretization (SID) [4]
  - Uniform Discretization (UD) [22]

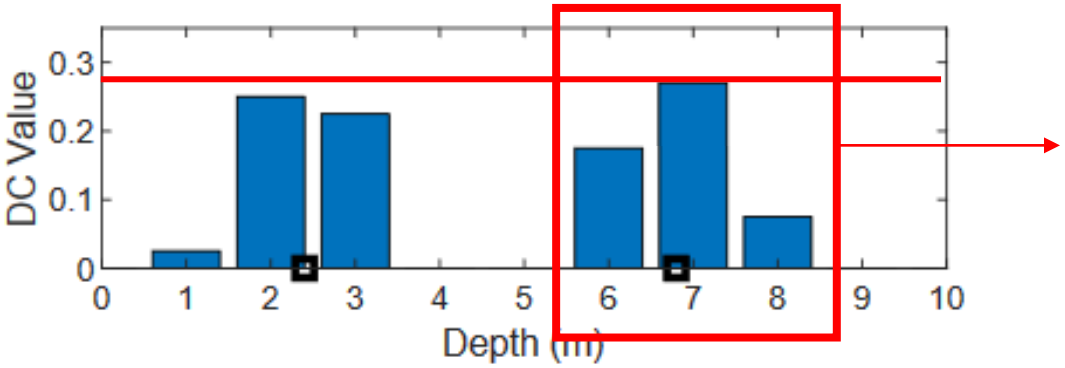
$$UD:\ t_i = d_L + (d_U - d_L) * i/K$$

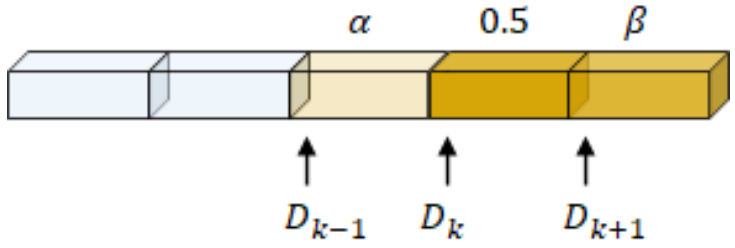
  - Spacing Increasing Discretization (SID) [4]

$$SID:\ t_i = e^{\log(d_L) + \frac{log(d_U - d_L) * i}{K}}$$



Depth Coefficient → Depth Value

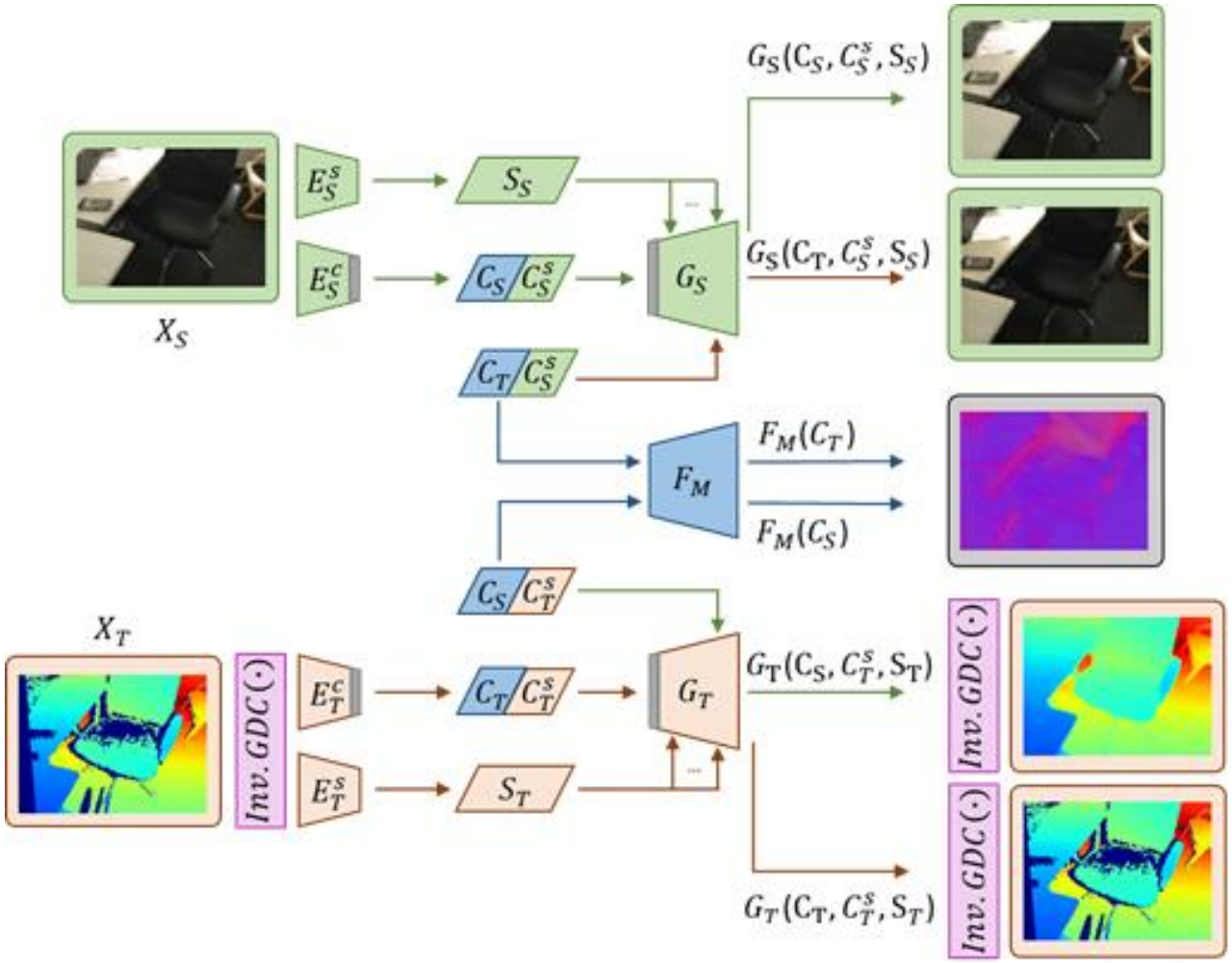




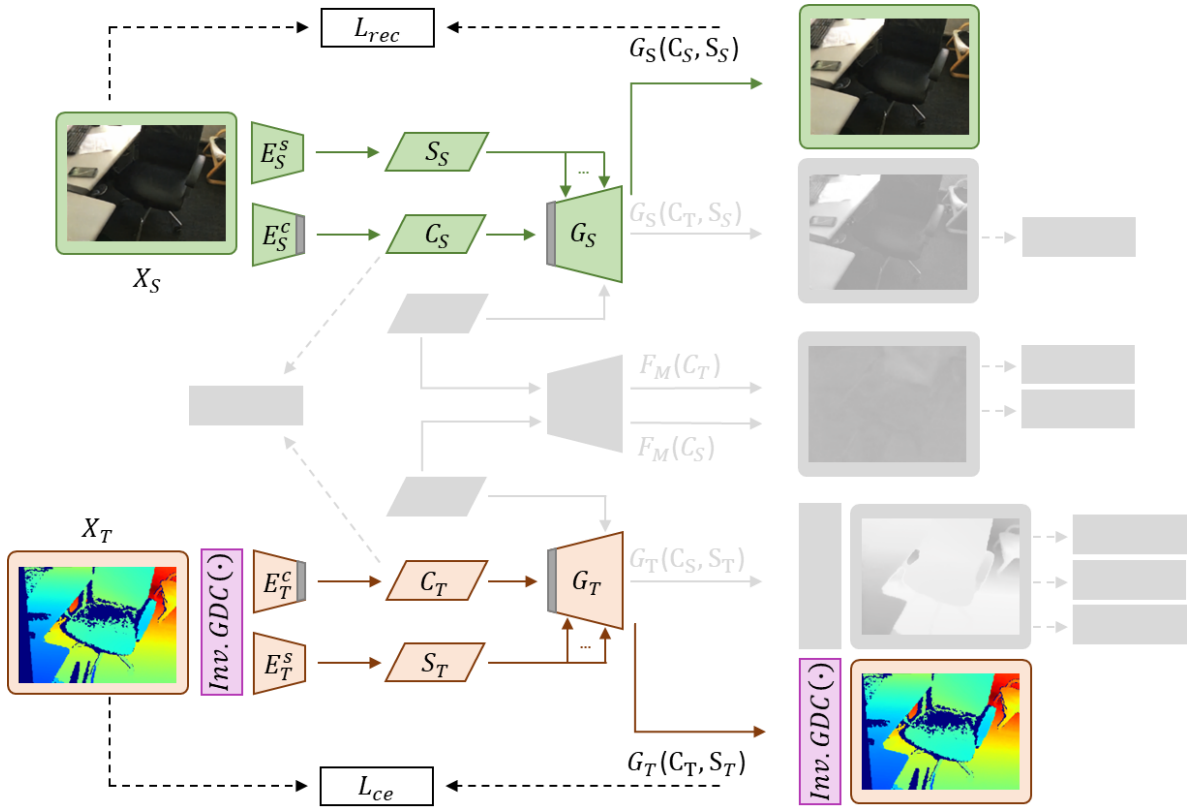
$$\hat{d}_i = \frac{\hat{c}_{i(k-1)} D_{(k-1)} + \hat{c}_{ik} D_k + \hat{c}_{i(k+1)} D_{(k+1)}}{\hat{c}_{i(k-1)} + \hat{c}_{ik} + \hat{c}_{i(k+1)}}$$

[4] Deep Ordinal Regression Network for Monocular Depth Estimation, CVPR, 2018
[22]Depth Coefficients for Depth Completion, arXiv, 2019

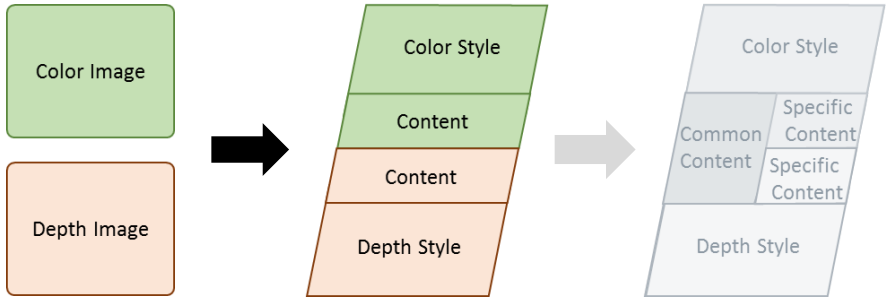
Network Architecture



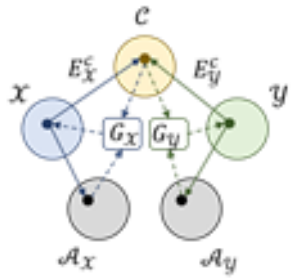
Network Architecture

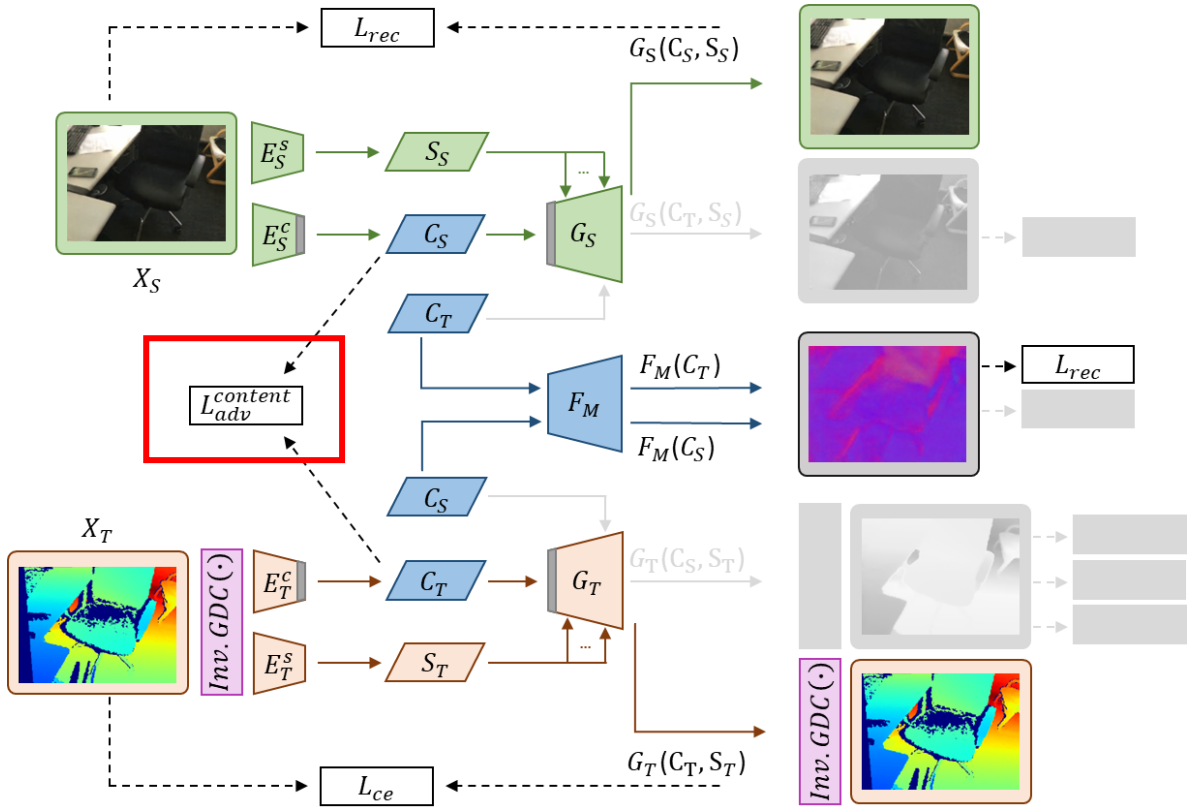


- **Disentangled Representation Learning**
- Domain Adaptation
- Feature exchange across domains



Network Architecture

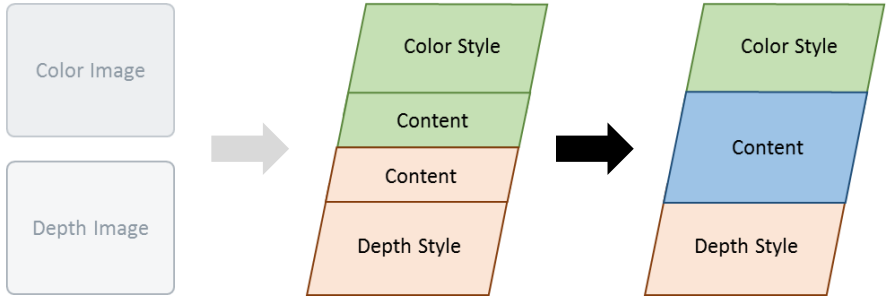




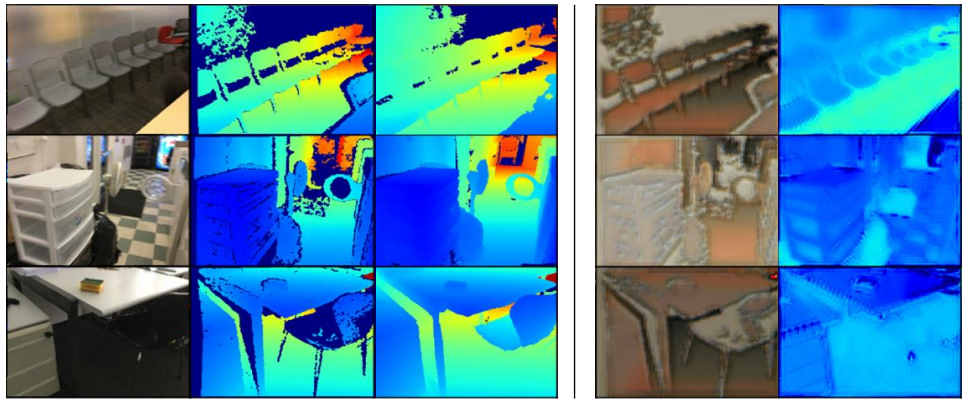
$$L_{adv}^{content}(D_M) = \mathbb{E}_{C_S}\left[\frac{1}{2}\log\big(D_M(C_S)\big) + \frac{1}{2}\log\big(1 - D_M(C_S)\big)\right]$$

$$+ \mathbb{E}_{C_T}\left[\frac{1}{2}\log\big(D_M(C_T)\big) + \frac{1}{2}\log\big(1 - D_M(C_T)\big)\right]$$

- Disentangled Representation Learning
- **Domain Adaptation**
- Feature exchange across domains

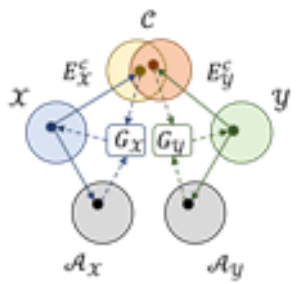


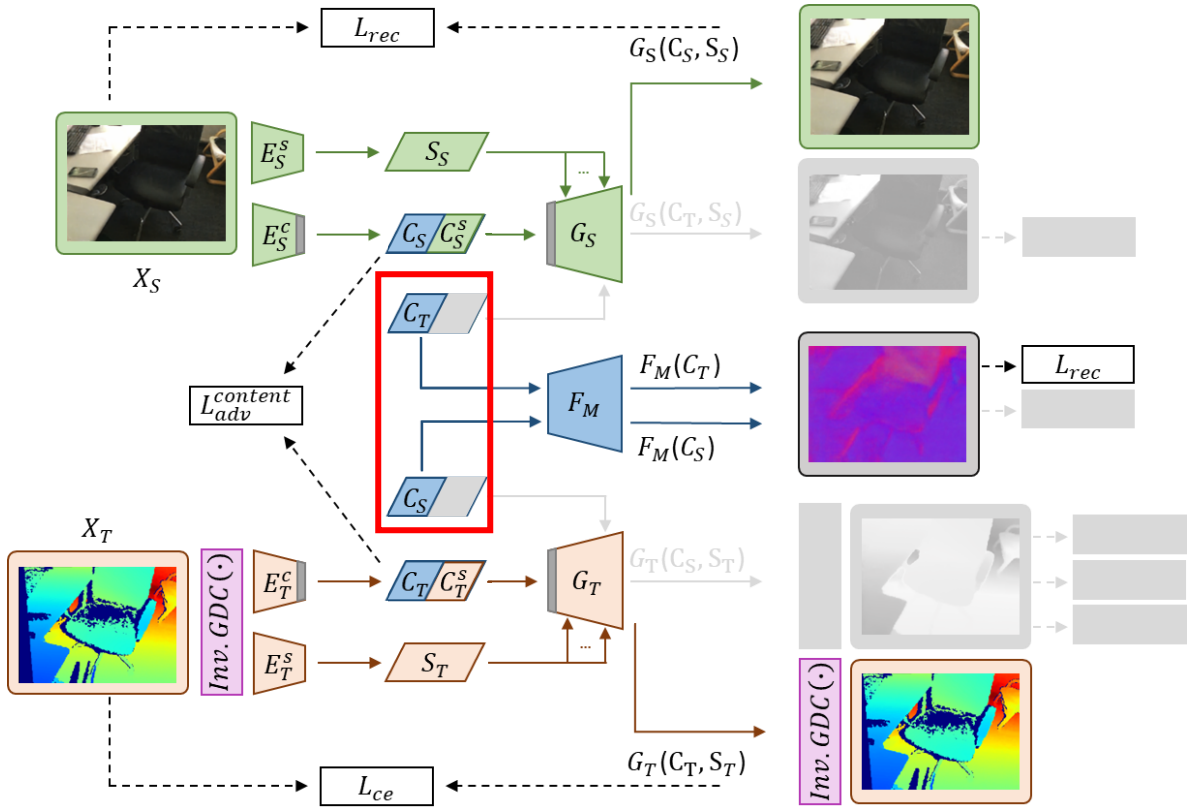
Fail if force the model to align content domain directly



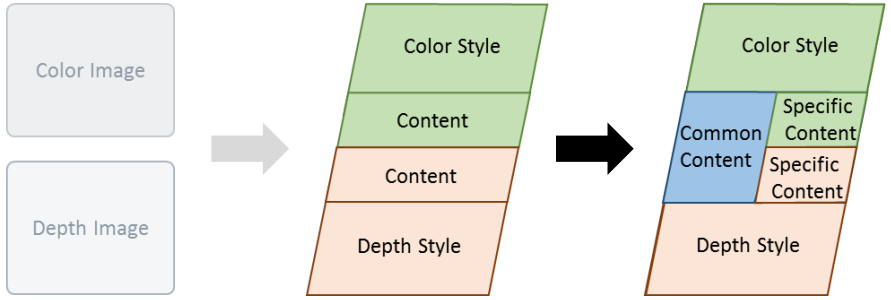
X_S X_T Ground truth $G_S(C_T, S_S)$ $G_T(C_S, S_T)$

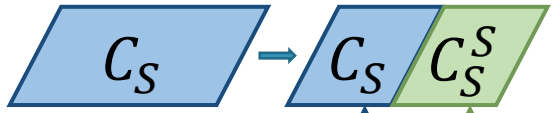
Network Architecture





- Disentangled Representation Learning
- **Domain Adaptation**
- Feature exchange across domains

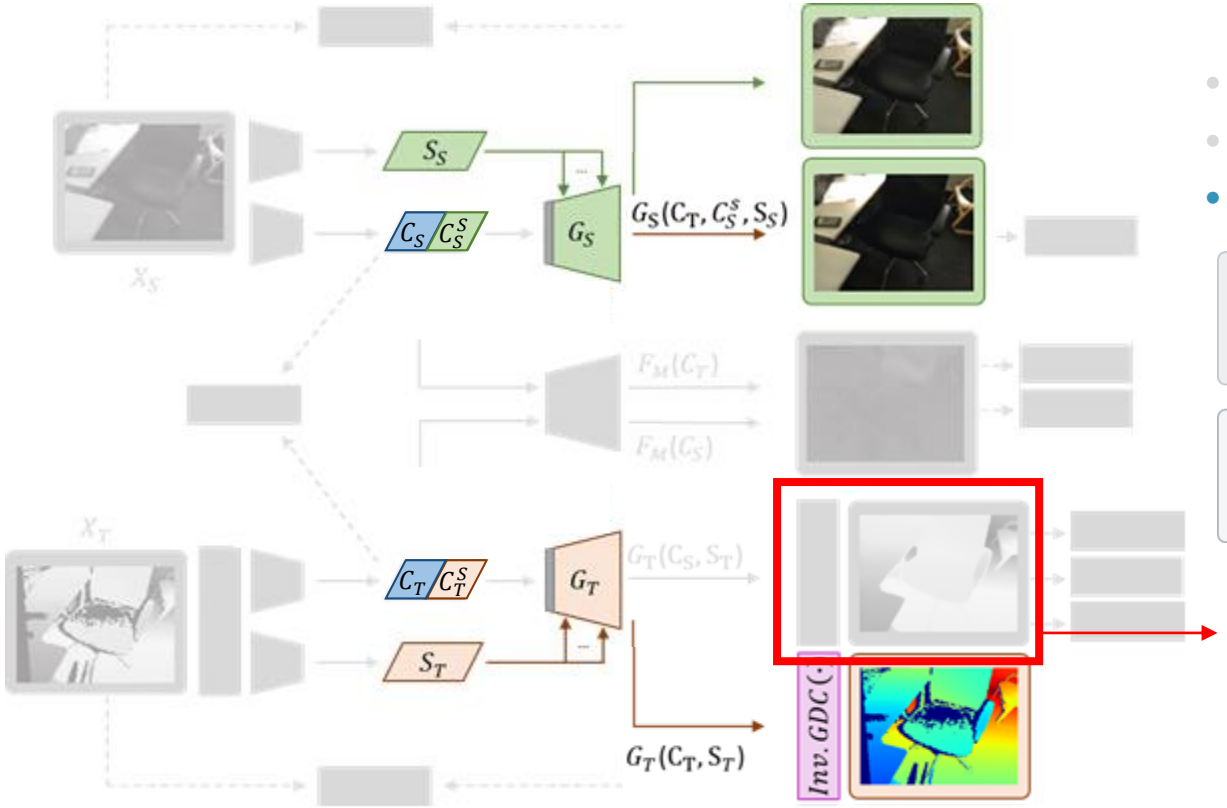




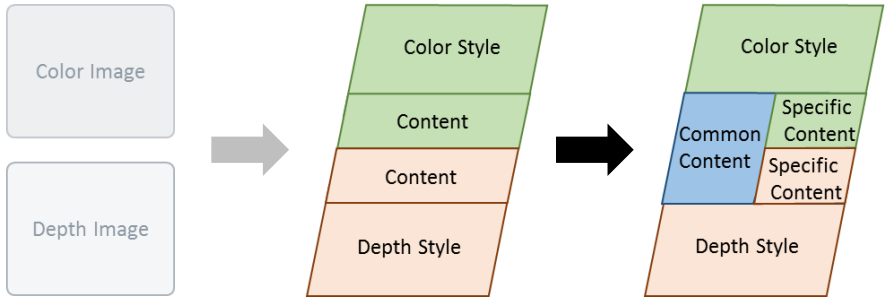
Learn the structural common content from both color and depth domain

The specific information for reconstruction

Network Architecture

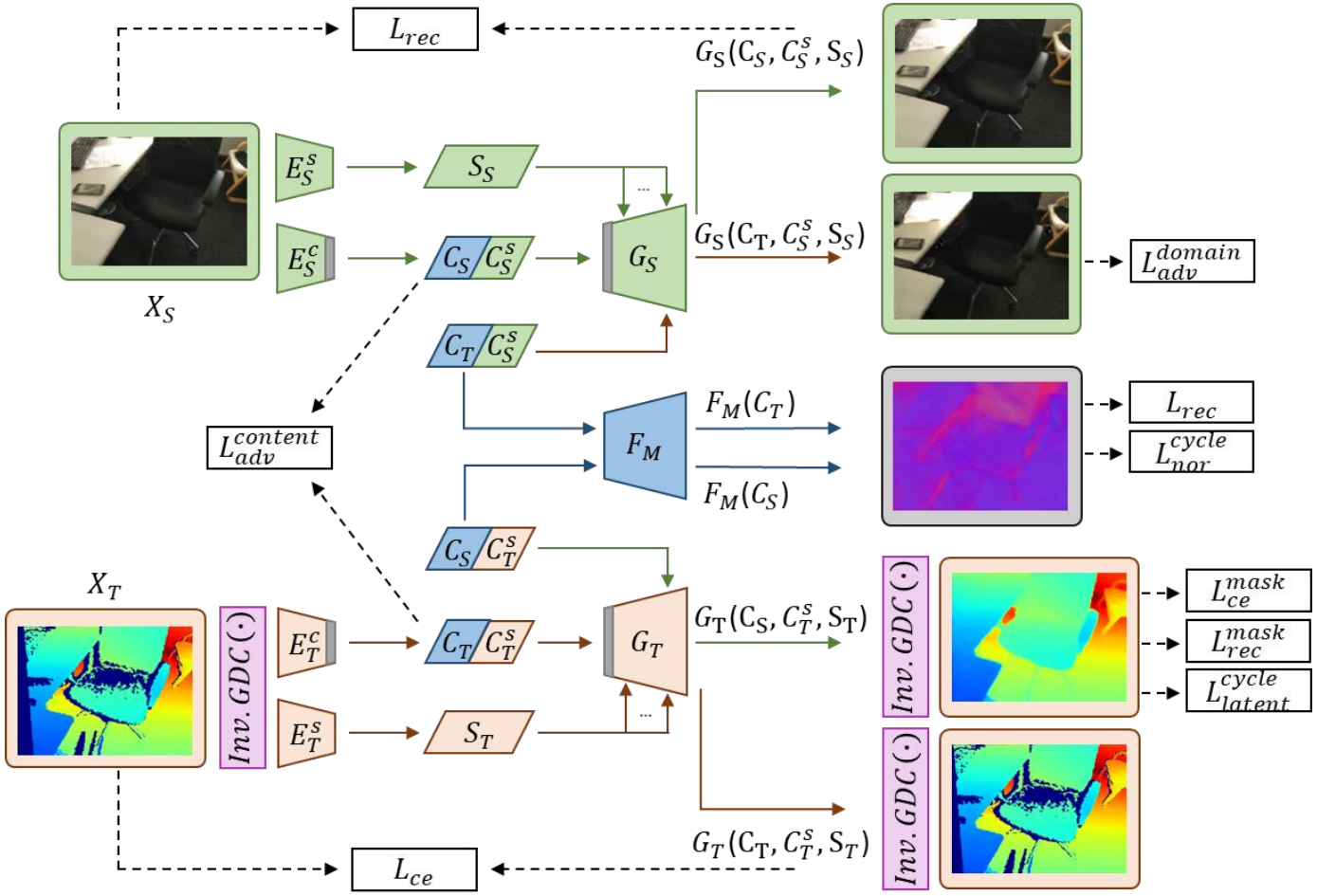


- Disentangled Representation Learning
- Domain Adaptation
- **Feature exchange across domains**

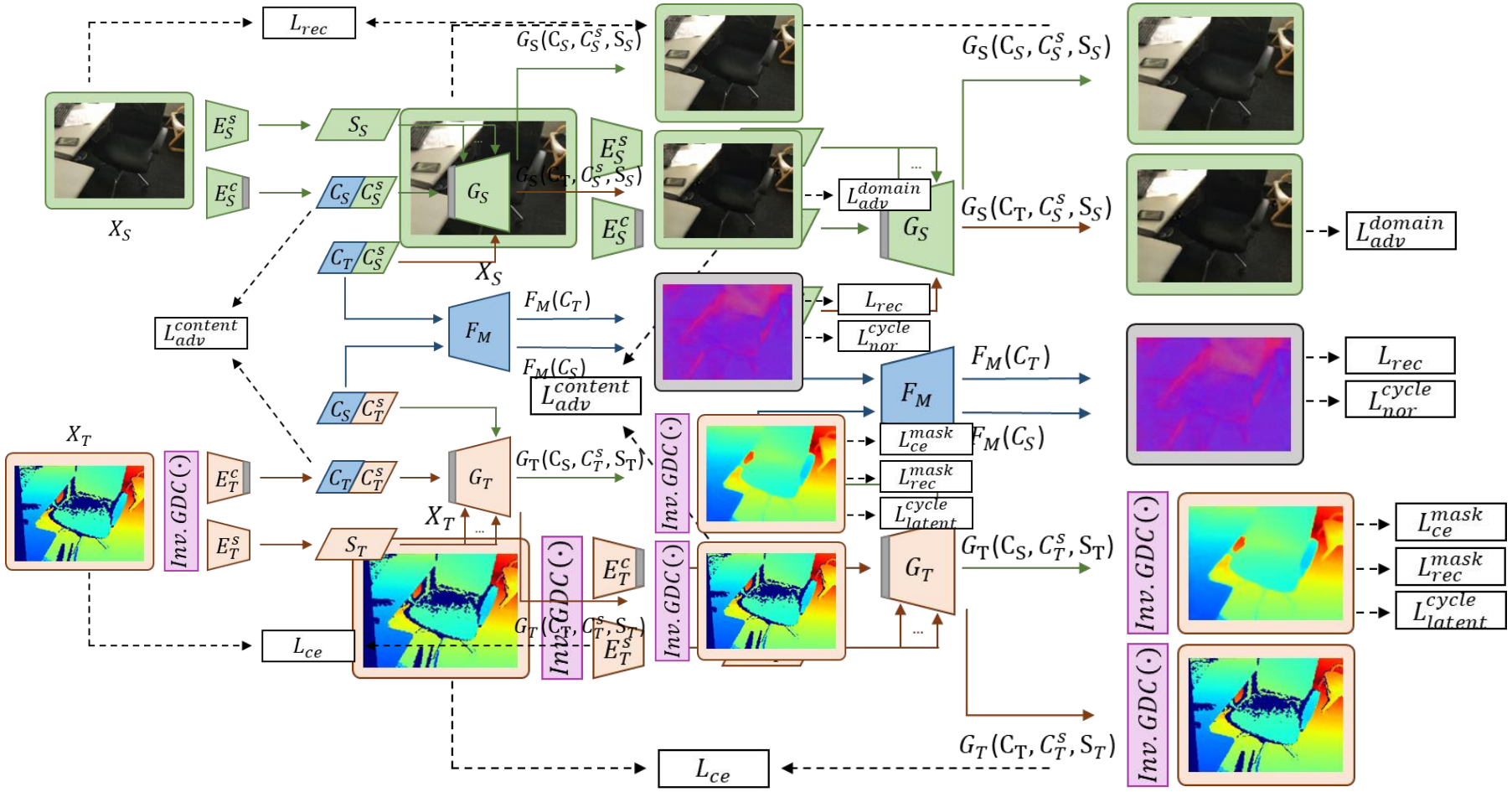


Final Output

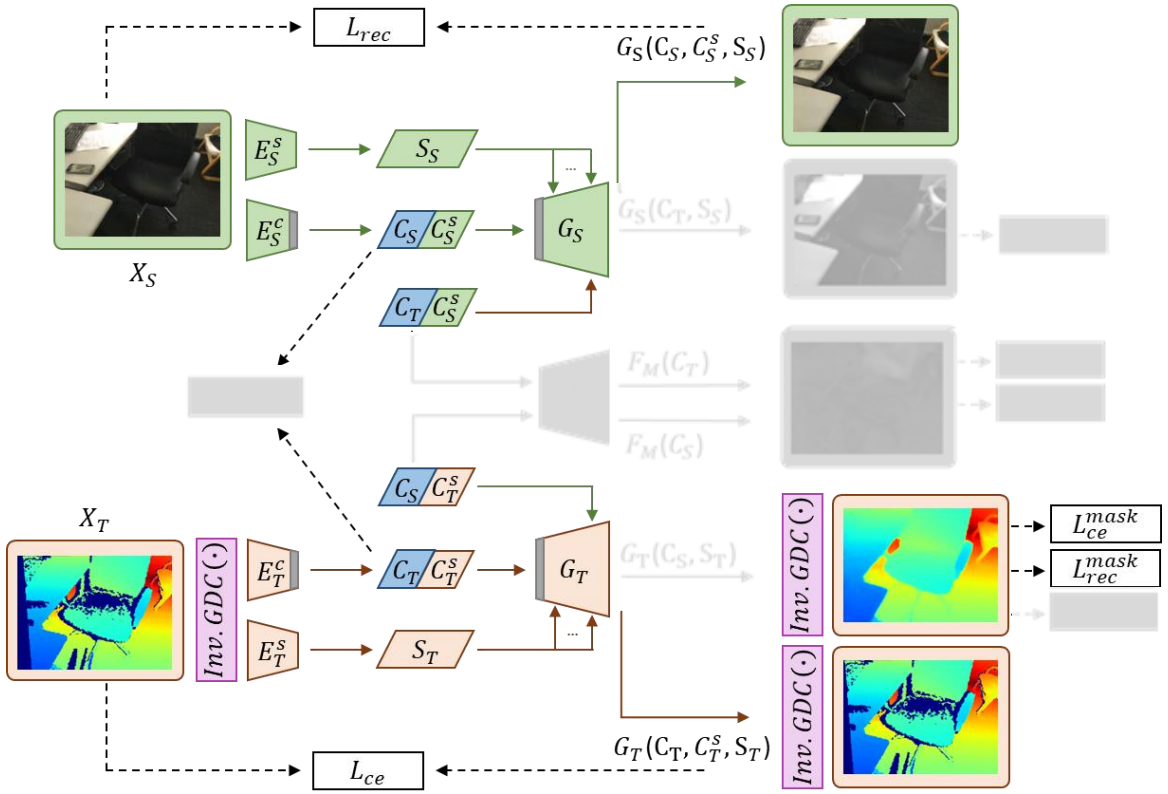
Criterion Design



Criterion Design



Criterion Design

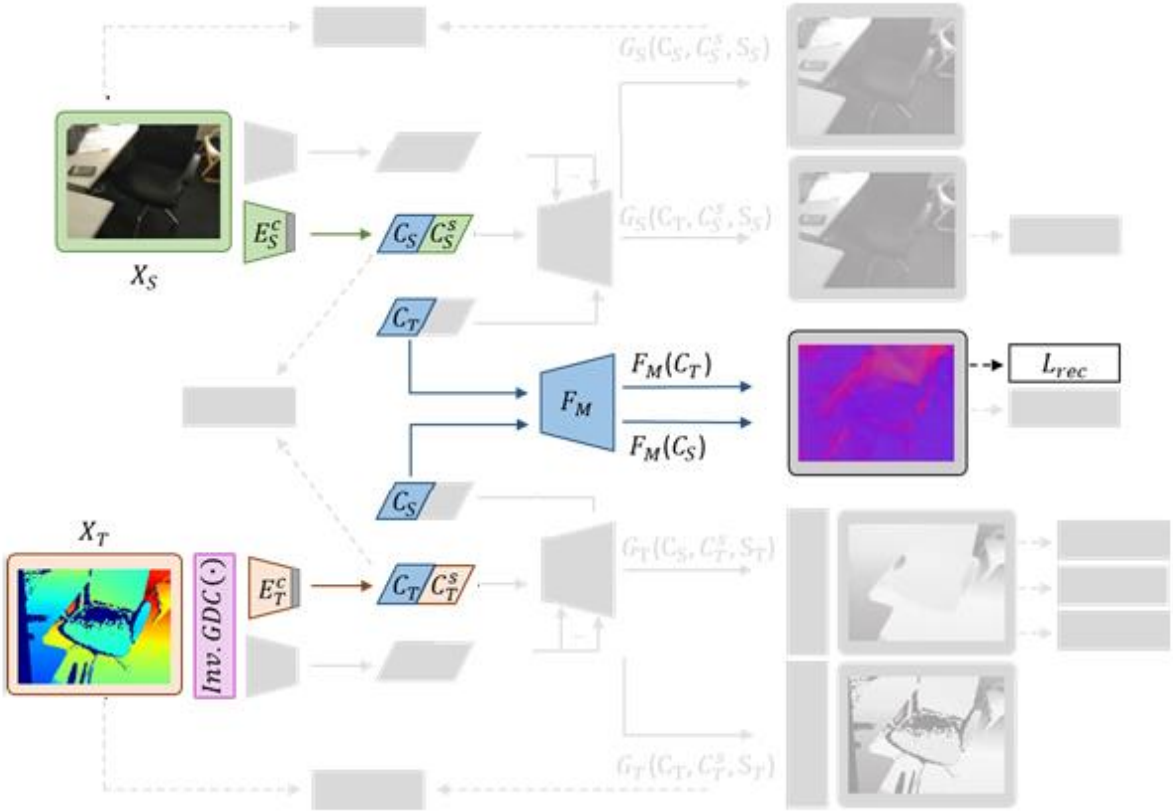


What kind of loss do we need?

- L_{rec}: $l_i^{rec}(m_i, d_i, \hat{d}_i) = -m_i(d_i - \hat{d}_i)^2$
- L_{ce}: $l_i^{ce}(m_i, c_{ij}, \hat{c}_{ij}) = -m_i \sum_{j=1}^{K} c_{ij} log\, \hat{c}_{ij}$

Reconstruction Loss

Criterion Design



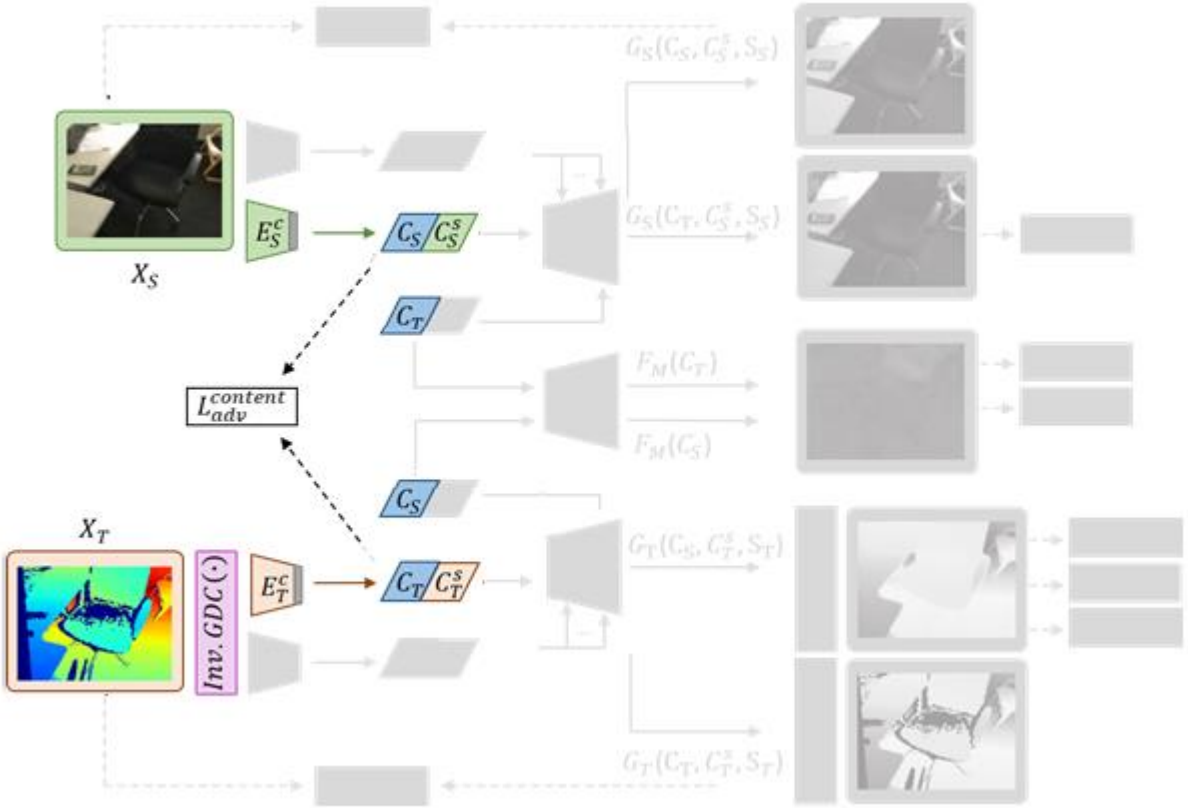
What kind of loss do we need?

- L_{rec}: $l_i^{rec}(m_i, d_i, \hat{d}_i) = -m_i(d_i - \hat{d}_i)^2$
- L_{ce}: $l_i^{ce}(m_i, c_{ij}, \hat{c}_{ij}) = -m_i \sum_{j=1}^{K} c_{ij} log\, \hat{c}_{ij}$

Reconstruction Loss

Surface Normal Loss

Criterion Design



What kind of loss do we need?

- L_{rec}: $l_i^{rec}(m_i, d_i, \hat{d}_i) = -m_i(d_i - \hat{d}_i)^2$
- L_{ce}: $l_i^{ce}(m_i, c_{ij}, \hat{c}_{ij}) = -m_i \sum_{j=1}^{K} c_{ij} \log \hat{c}_{ij}$

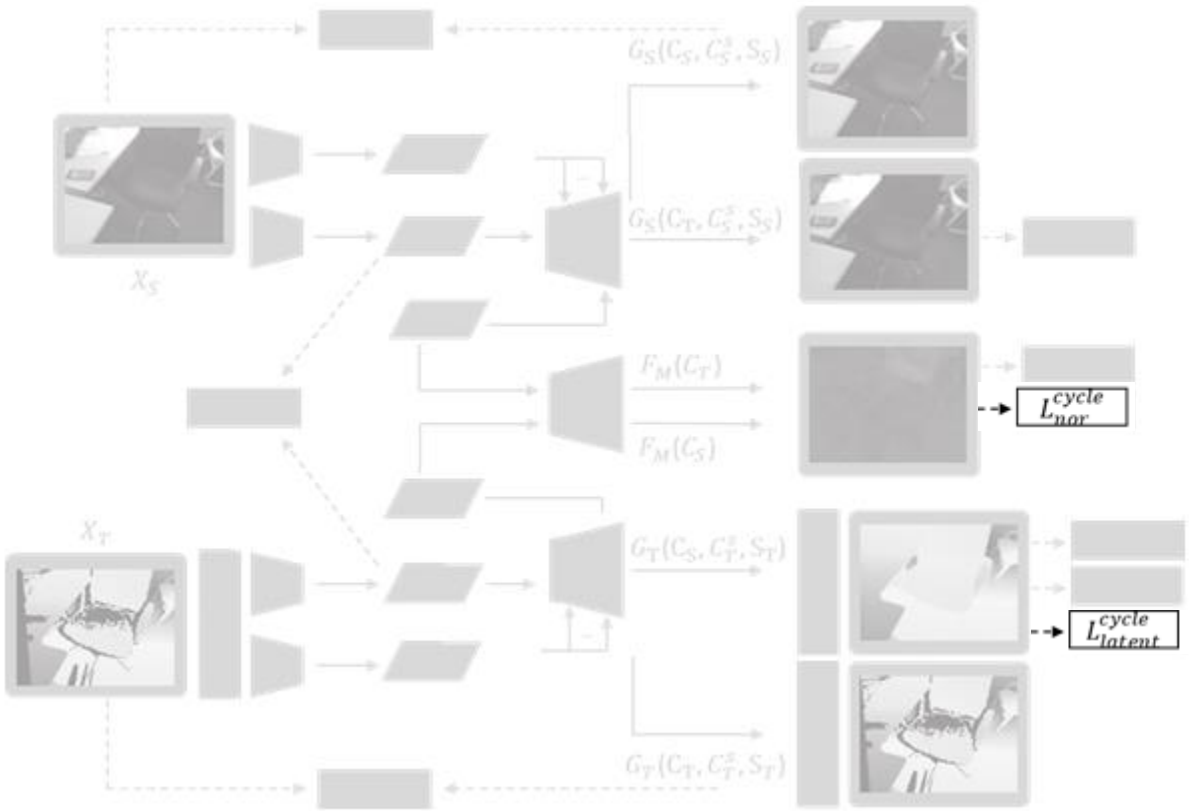
Reconstruction Loss

Surface Normal Loss

Adversarial Loss

$$L_{adv}^{content}(D_M) = \mathbb{E}_{C_S}\left[\frac{1}{2}\log\big(D_M(C_S)\big) + \frac{1}{2}\log\big(1 - D_M(C_S)\big)\right] + \mathbb{E}_{C_T}\left[\frac{1}{2}\log\big(D_M(C_T)\big) + \frac{1}{2}\log\big(1 - D_M(C_T)\big)\right]$$

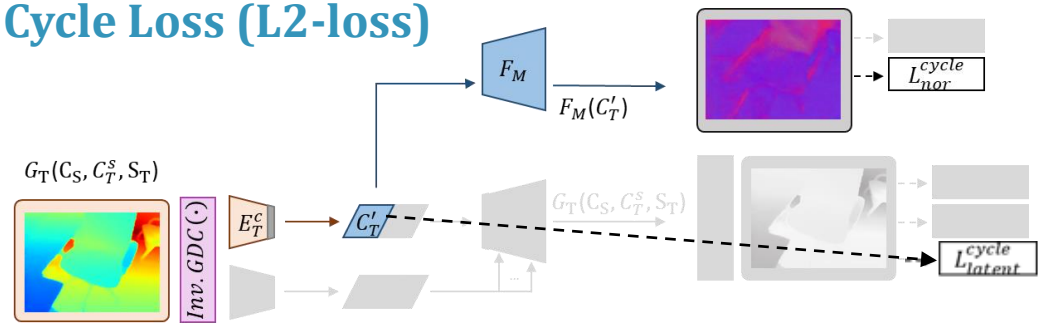
Criterion Design



What kind of loss do we need?

- L_{rec}: $l_i^{rec}(m_i, d_i, \hat{d}_i) = -m_i(d_i - \hat{d}_i)^2$
- L_{ce}: $l_i^{ce}(m_i, c_{ij}, \hat{c}_{ij}) = -m_i \sum_{j=1}^{K} c_{ij} log\ \hat{c}_{ij}$

Reconstruction Loss
Surface Normal Loss
Adversarial Loss
Cycle Loss (L2-loss)

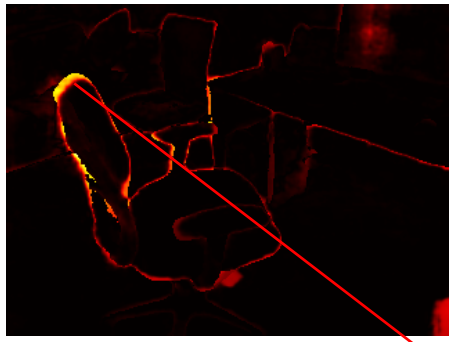


Experiment

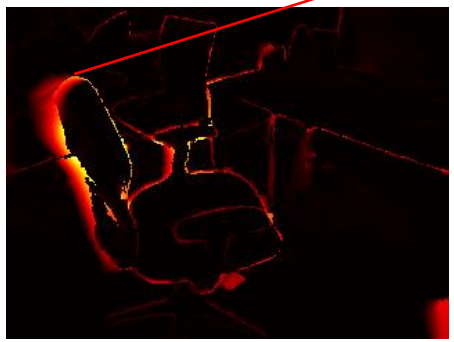
Results

Comparison

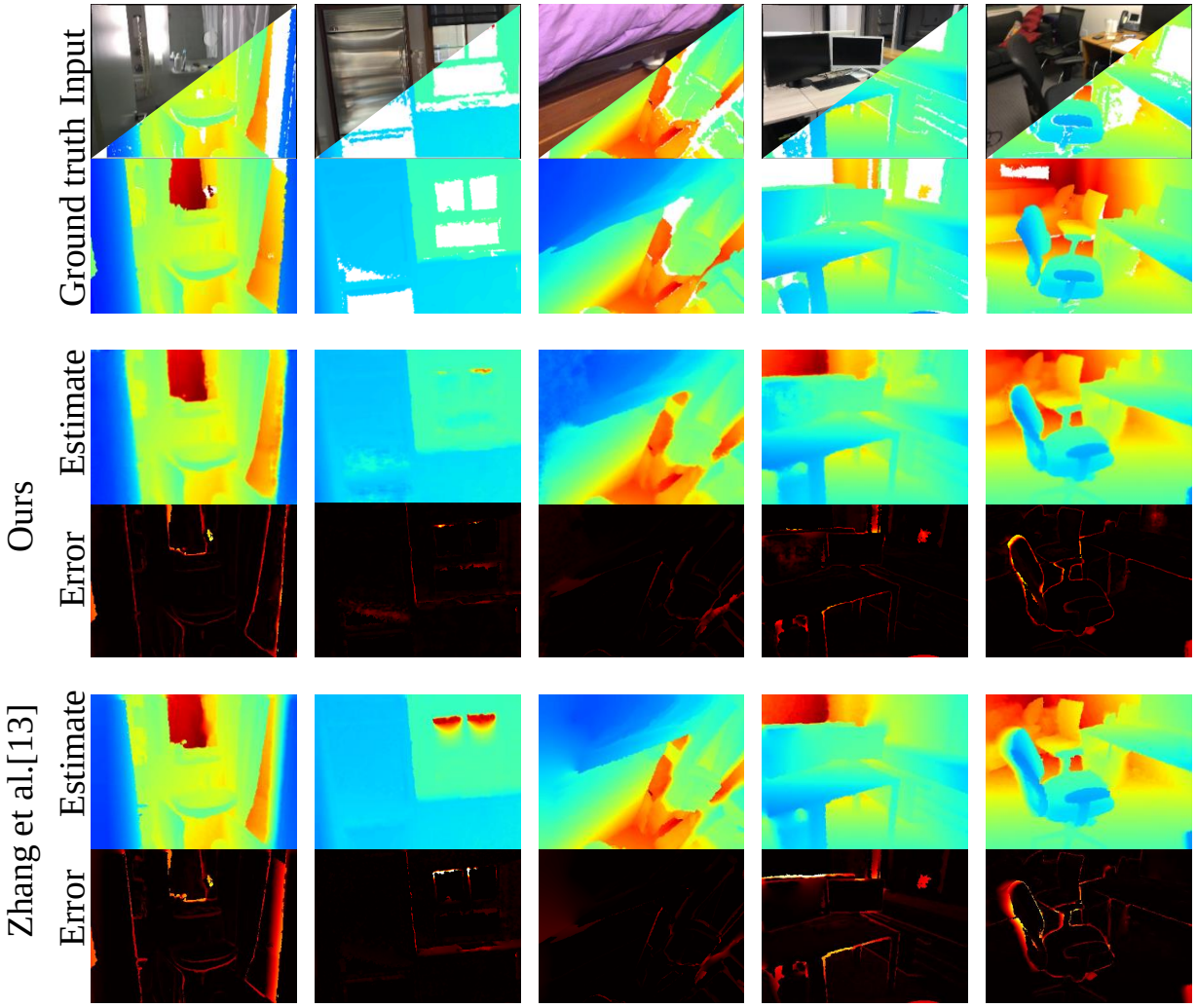
Ours



Zhang et al.[13]



Much Less
Mixed Depth Pixel



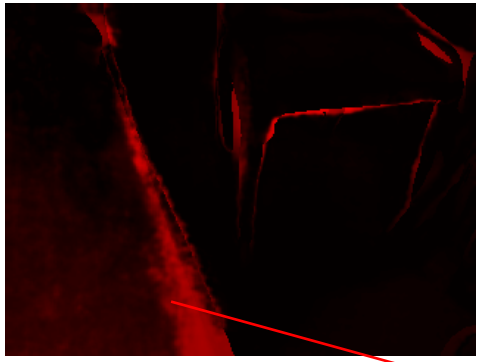
Example depth completion results on ScanNet test set.

[13] Deep Depth Completion of a Single RGB-D Image, CVPR, 2018

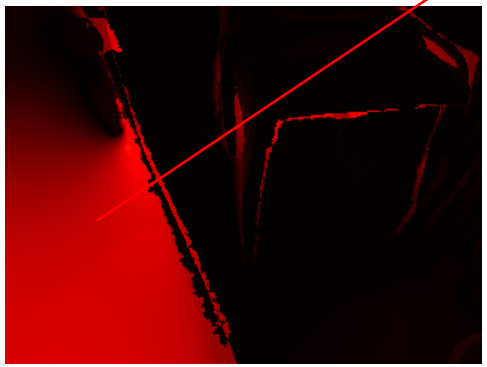
Results

Comparison

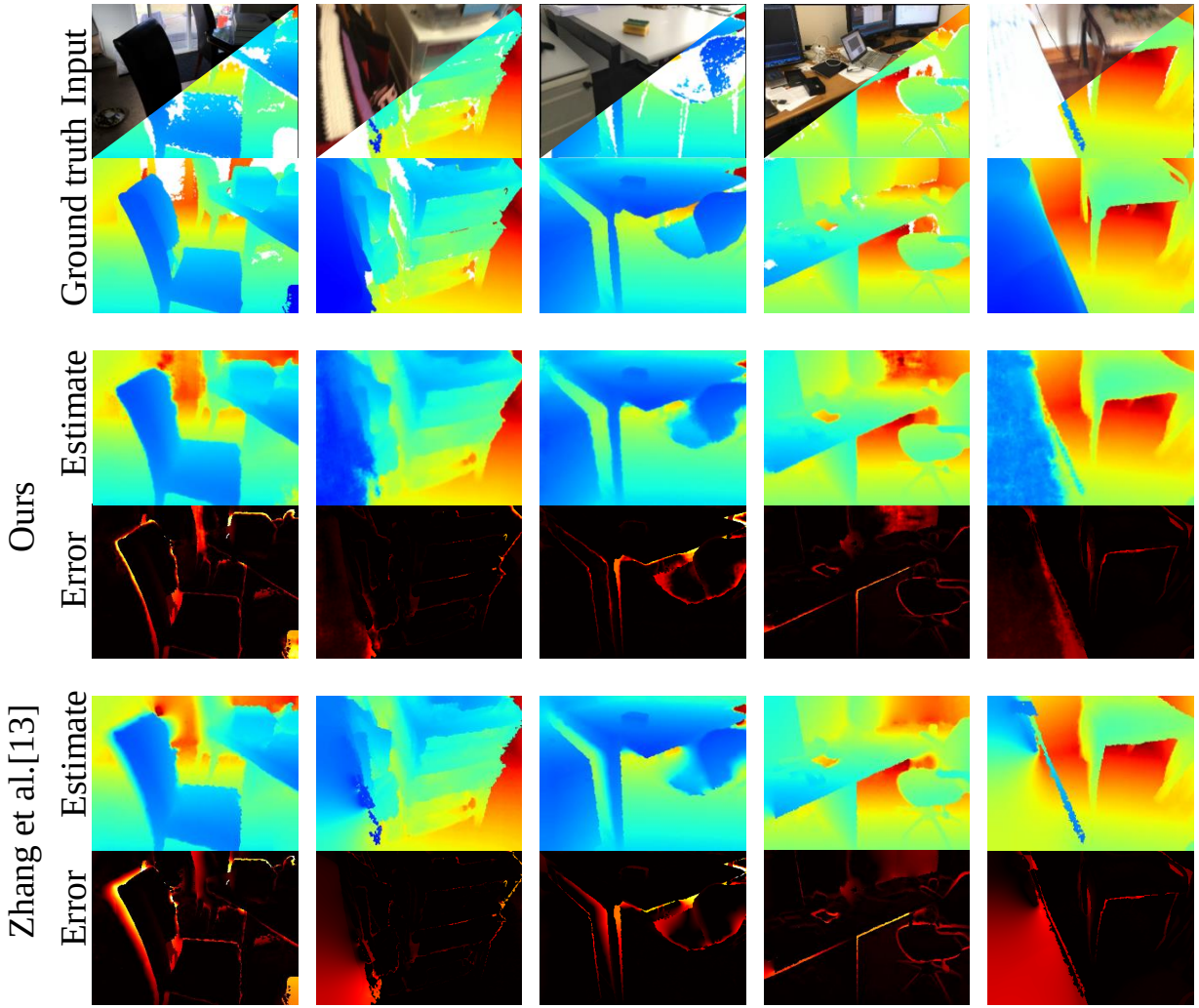
Ours



Zhang et al.[13]



Much Less
Spatial Scale Offset

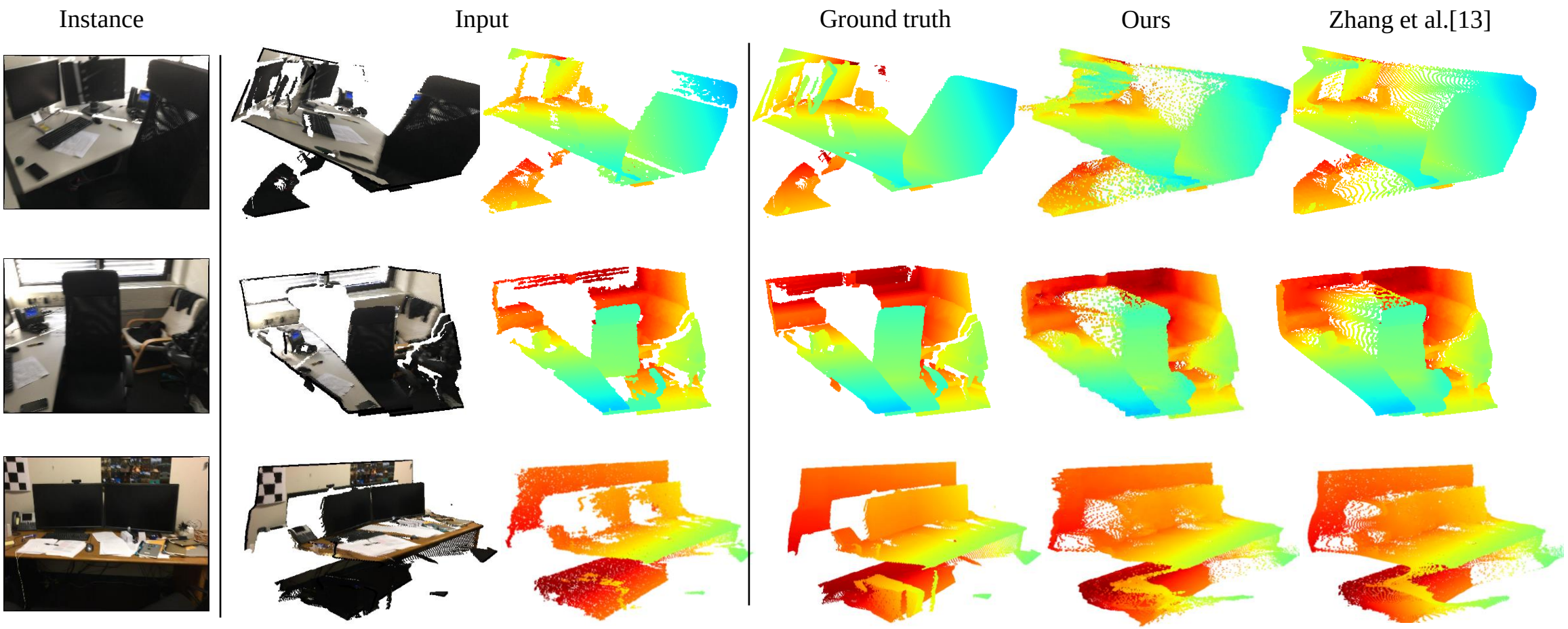


Example depth completion results on ScanNet test set.

[13] Deep Depth Completion of a Single RGB-D Image, CVPR, 2018

Results

Comparison



Point cloud visualization of our method and other comparisons.
We convert the completed depth into point cloud.

[13] Deep Depth Completion of a Single RGB-D Image, CVPR, 2018

Results

Comparison

Obs.	Method	REL ↓	RMSE↓	1.25↑	1.25^2 ↑	1.25^3 ↑
B	Zhang et al.[13]	0.0100	0.0155	0.9213	0.9588	0.9764
	Ours(GDC)	**0.0085**	**0.0132**	**0.9247**	**0.9621**	**0.9794**
Y	Zhang et al.[13]	0.0076	0.0117	0.9588	0.9757	0.9856
	Ours(GDC)	**0.0063**	**0.0096**	**0.9617**	**0.9786**	**0.9877**
N	Zhang et al.[13]	0.0408	0.0637	0.8113	0.9092	0.9492
	Ours(GDC)	**0.0386**	**0.0590**	**0.8160**	**0.9134**	**0.9551**

Comparison against state-of-the-art algorithm on ScanNet dataset.
(B: GDT>0, Y: GDT>0 & RAW>0, N: GDT>0 & RAW=0)
Best result show in blue.

[13] Deep Depth Completion of a Single RGB-D Image, CVPR, 2018

Thank You For Listening