iDisc: Internal Discretization for Monocular Depth Estimation

Luigi Piccinelli, Christos Sakaridis, Fisher Yu

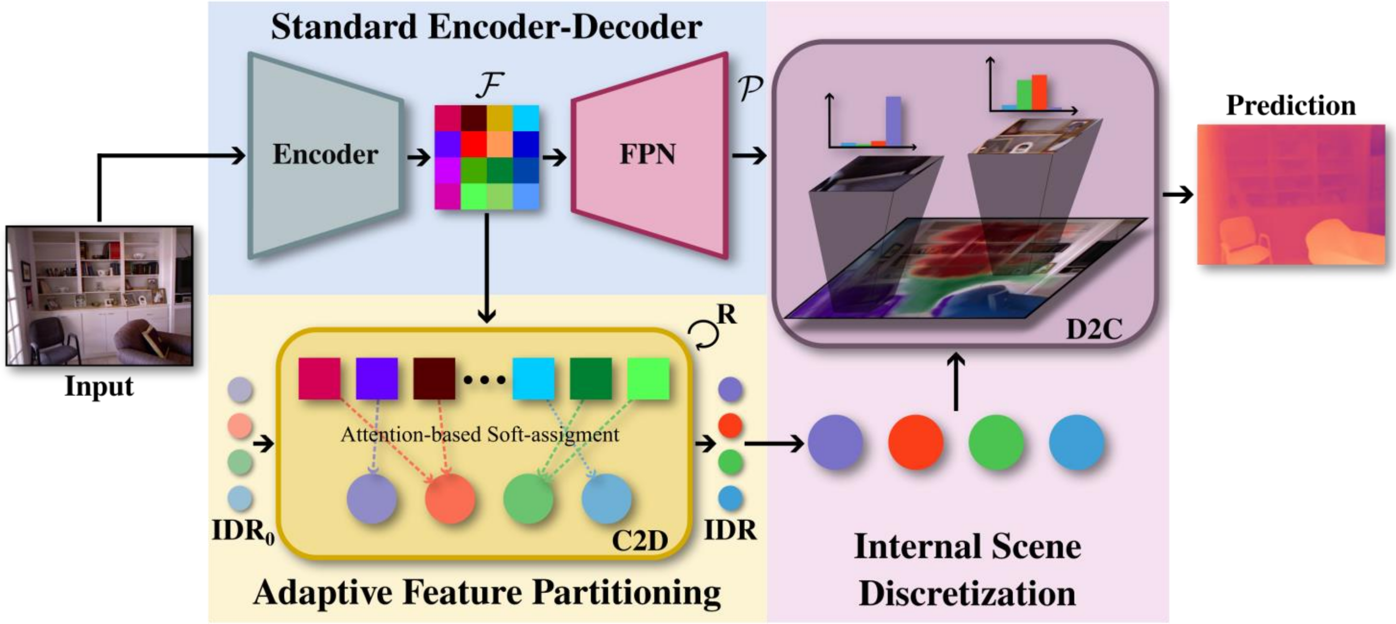
Poster THU-PM-083

Project page: vis.xyz/pub/idisc

Code and models: github.com/SysCV/idisc

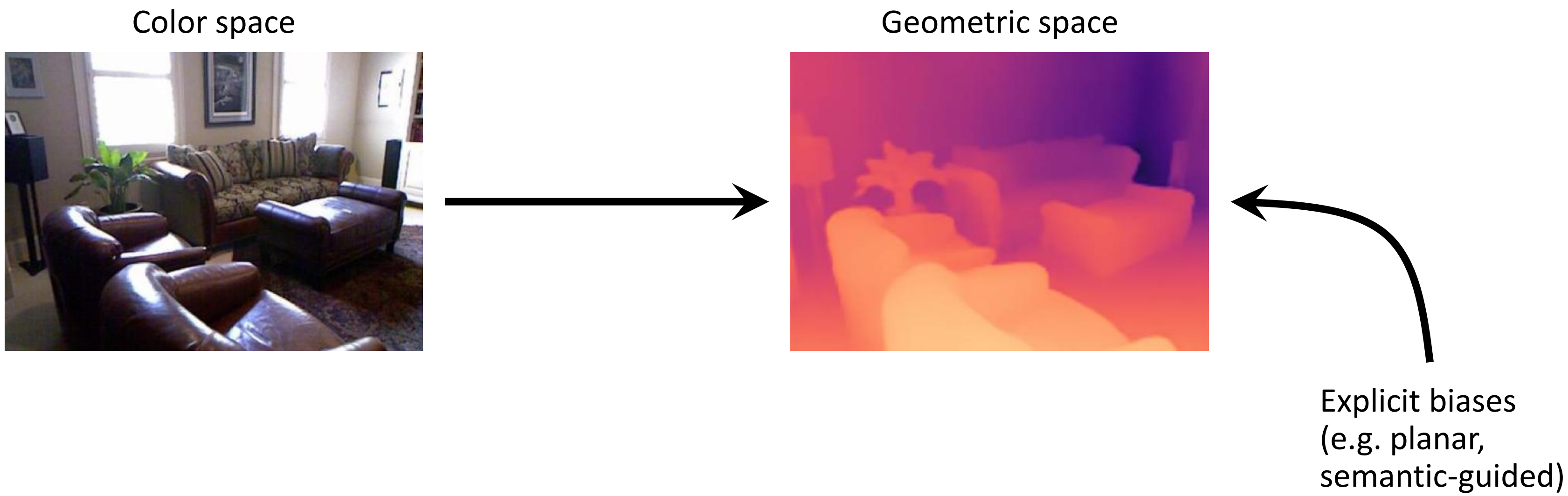
Overview iDisc

- Lift any handcrafted bias imposed on the scene representation.
- One assumption only: scene is a discrete set of high-level concepts.
- iDisc meta-learns the best internal representations.



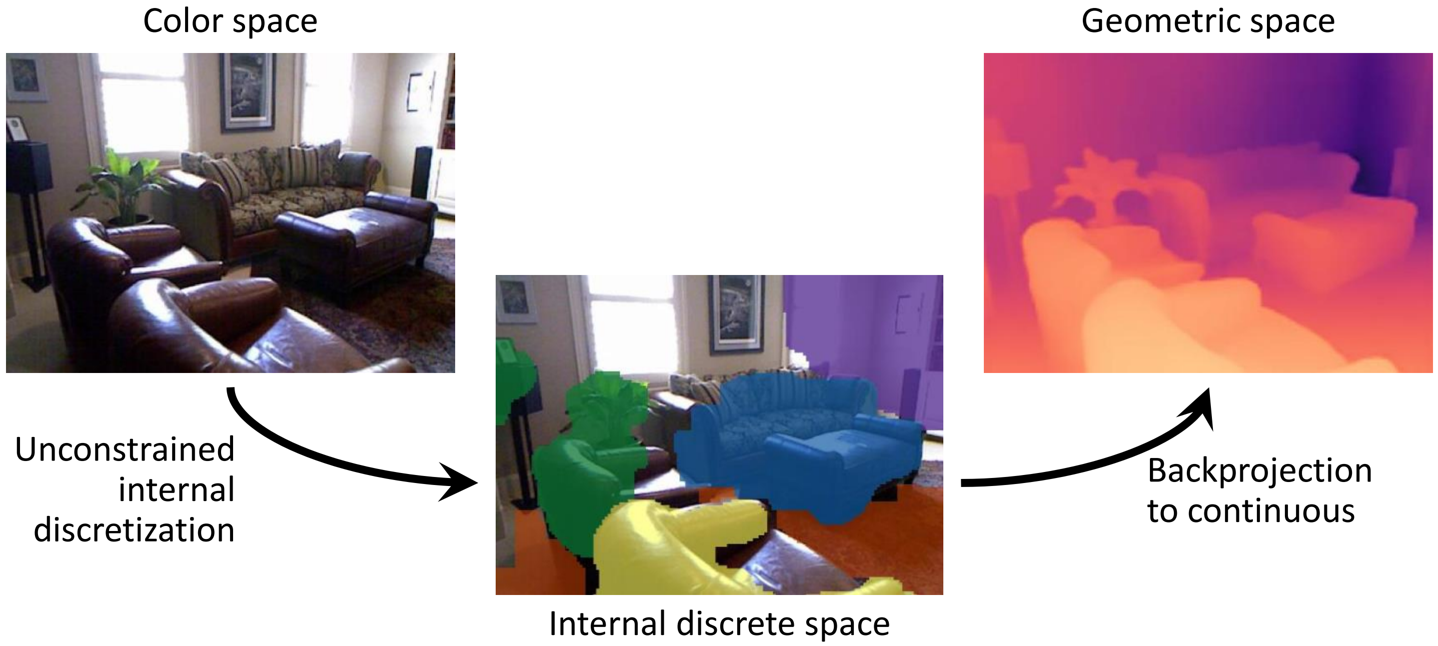
Monocular and biases

- Ill-posed problem, priors are needed.
- Typically, the scene representation is handcraftedly biased.
- Can it learn how to generate appropriate “priors” for the given input?



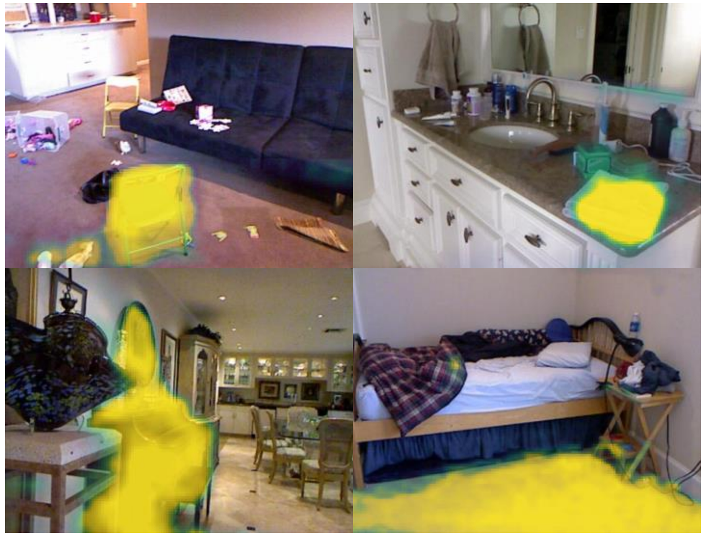
Monocular and biases

- Ill-posed problem, priors are needed.
- Typically, the scene representation is handcraftedly biased.
- Can it learn how to generate appropriate “priors” for the given input?

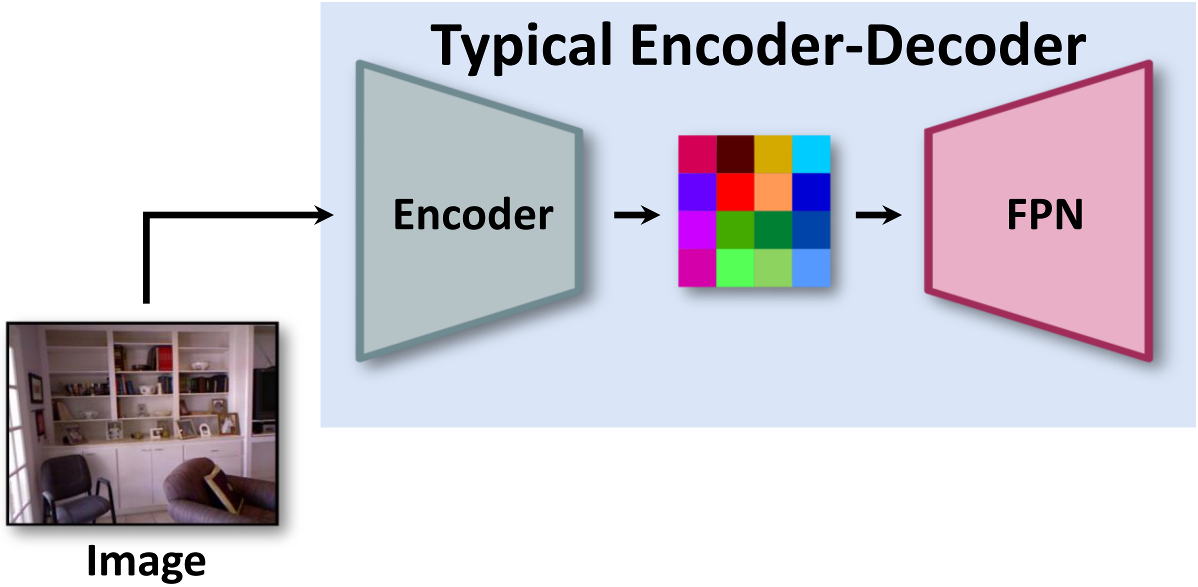


What is a concept?

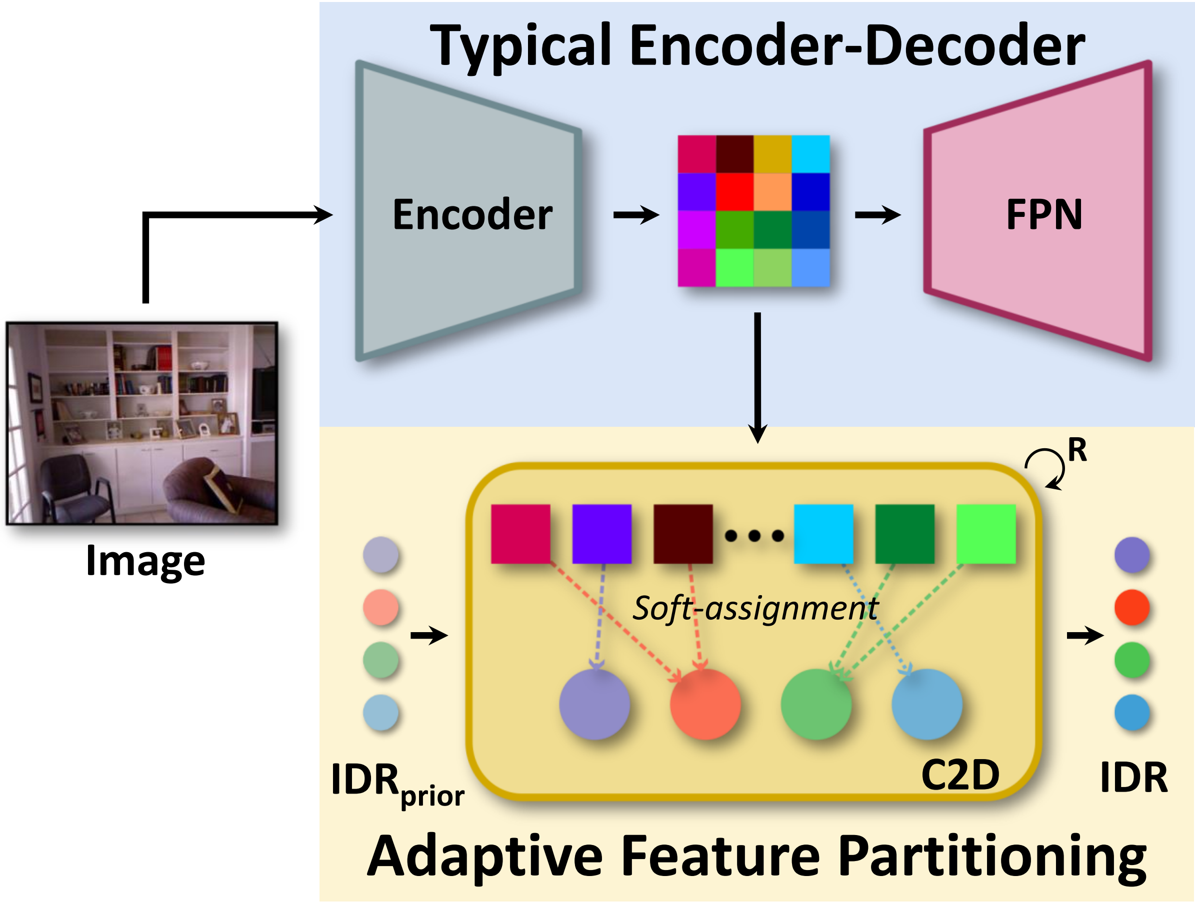
- Set of high-level structures deemed appropriate to describe the scene.
- Internal discrete representations learned without any supervision.



Architecture



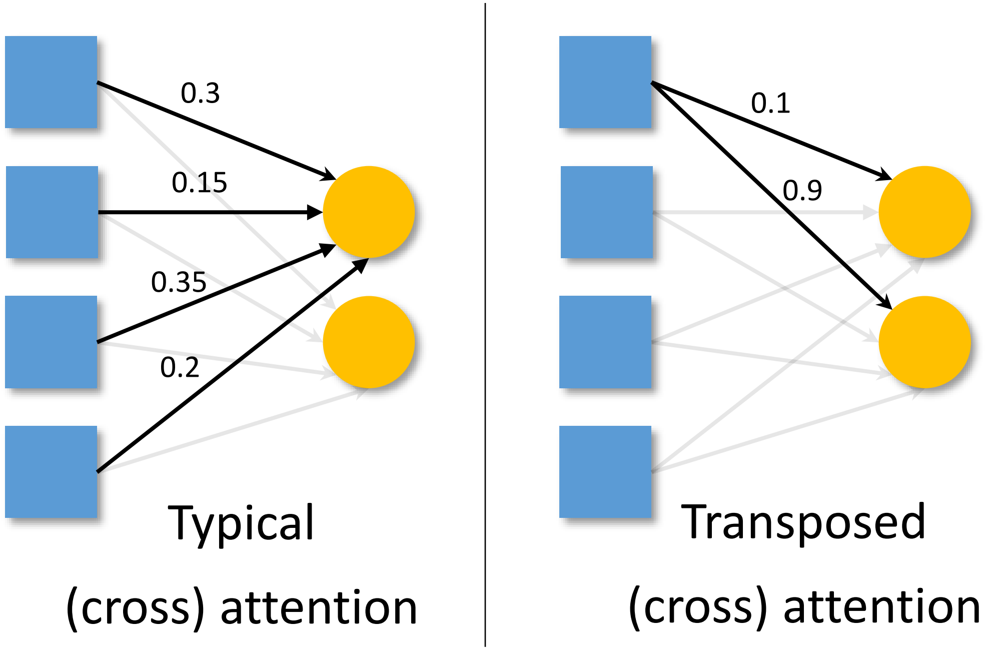
Architecture



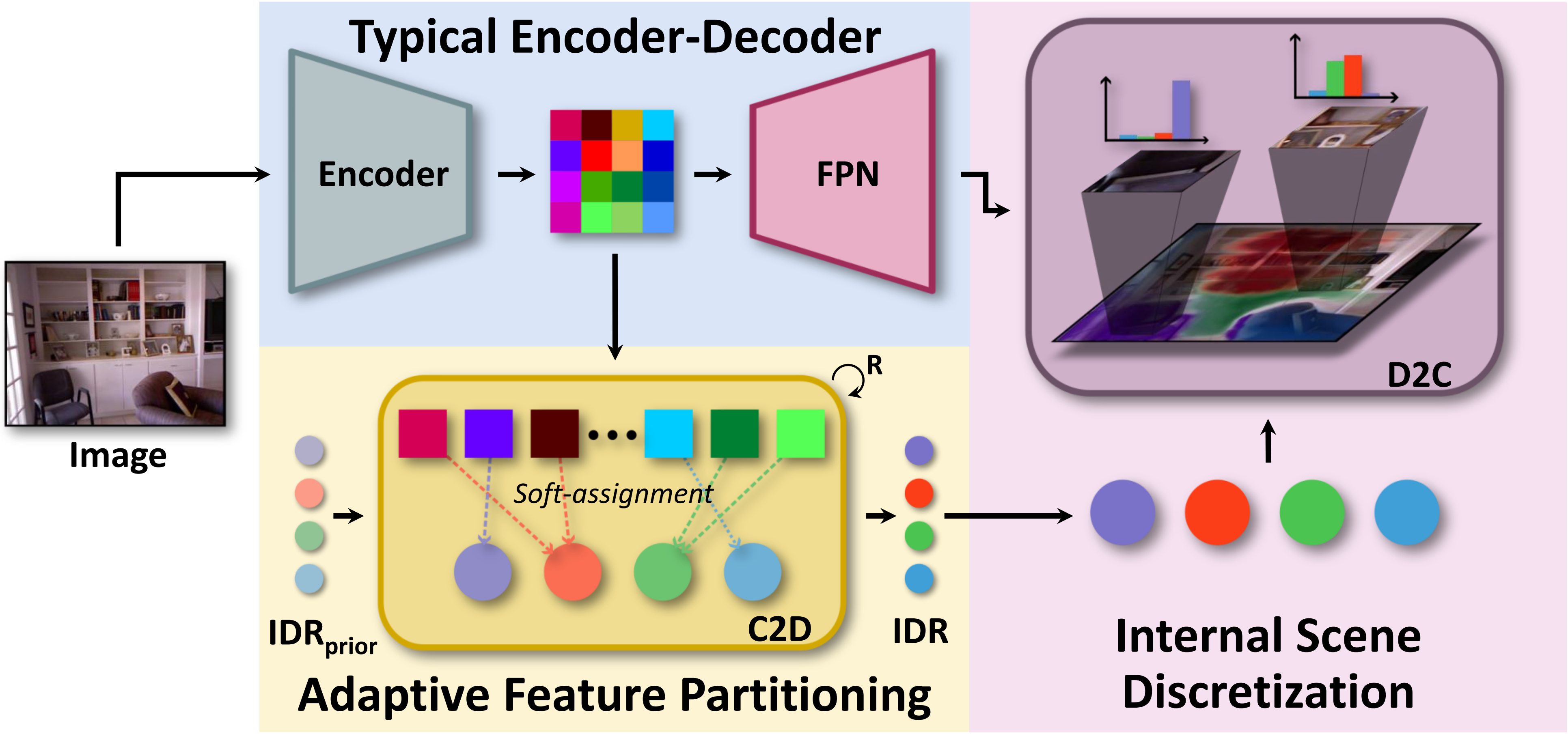
$$\mathbf{Q}_{i+1} = \left[\mathrm{softmax}(\mathbf{K}\mathbf{Q}_i^T)\right]^T \mathbf{V}$$
$$\mathbf{Q}_0 = \mathbf{IDR}_{\mathbf{prior}}$$
$$\mathbf{IDR} = \mathbf{Q}_R$$

Partitioning

Adaptive



Architecture



$$\mathbf{D}_{i+1} = \mathrm{softmax}(\mathbf{Q}_i\mathbf{K}_i^T)\mathbf{V}_i + \mathbf{D}_i, \quad i \in 0..N$$

$$\mathbf{Q}_i = f_{Q_i}(\mathbf{P})$$

$$\mathbf{K}_i = f_{K_i}(\mathbf{IDR})$$

$$\mathbf{V}_i = f_{V_i}(\mathbf{IDR})$$

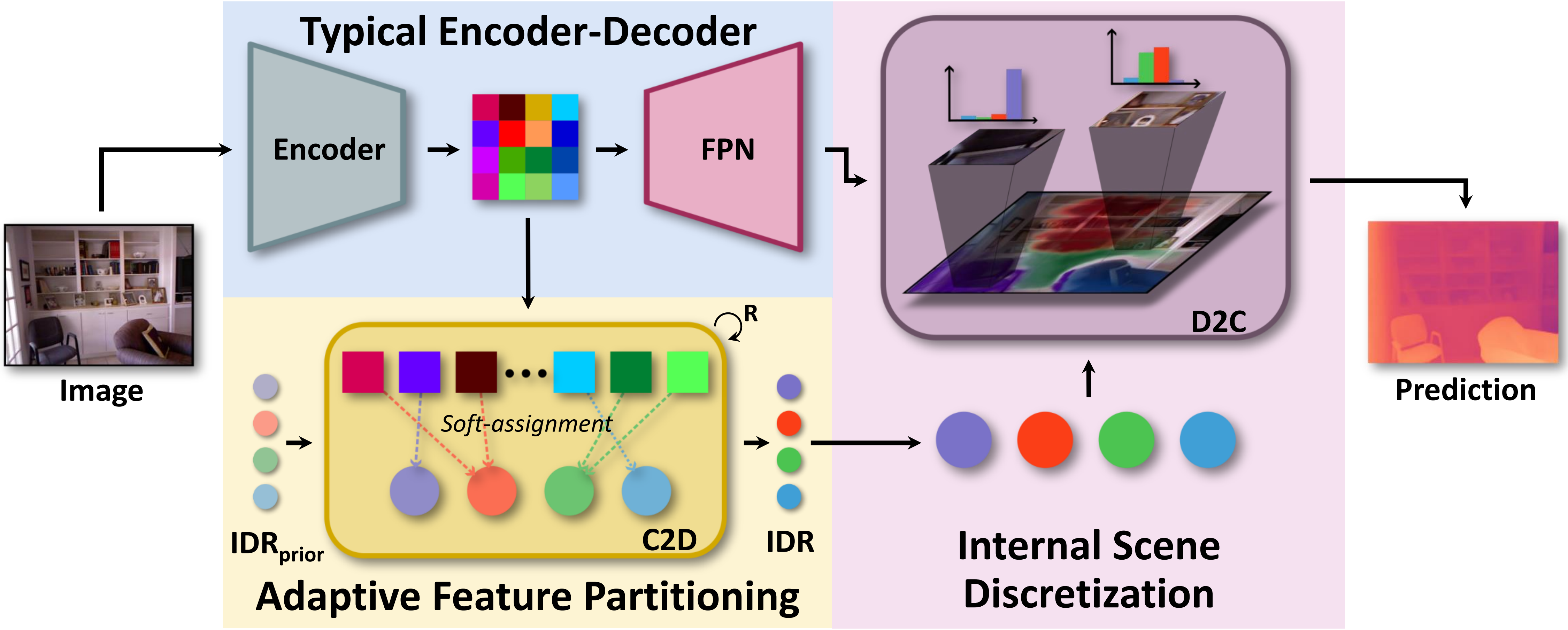
Degenerates to standard depth discretization if:

$$\mathbf{Q} = f_Q(\mathbf{P}), \ \mathbf{F} = \mathbf{K}||\mathbf{v}$$

$$\mathbf{D}_0 = \emptyset, \ N = 0, \ R = 0$$

$$\Rightarrow \mathbf{D} = \mathrm{softmax}(\mathbf{Q}\mathbf{K}^T)\mathbf{v}$$

Architecture



Quantitative results (common benchmarks)

Table 1. NYU-Depth v2 official test set results.

Method	δ_1 ↑	RMS ↓	A.Rel ↓
BTS	0.964	2.459	0.057
AdaBins	0.964	2.360	0.058
DPT	0.965	2.315	0.059
NeWCRF	0.974	2.129	0.052
iDisc	**0.977**	**2.067**	**0.050**

Quantitative results (common benchmarks)

Table 2. KITTI-Eigen split validation set results.

Method	$\delta_1 \uparrow$	RMS $\downarrow$	A.Rel $\downarrow$
BTS	0.885	0.392	0.110
AdaBins	0.903	0.364	0.103
DPT	0.904	0.357	0.110
NeWCRF	0.922	0.334	0.095
iDisc	**0.940**	**0.313**	**0.086**

Table 3. KITTI official online benchmark results.

Method	$SI_{log} \downarrow$	iRMS $\downarrow$	A.Rel $\downarrow$
ViP-DeepLab	10.80	11.77	0.089
NeWCRF	10.39	11.03	0.084
PixelFormer	10.29	10.84	0.082
iDisc	**9.89**	**10.73**	**0.081**

Quantitative results (proposed benchmarks)

Table 4. Results on Argoverse1.1 proposed split.

Method	$\delta_1 \uparrow$	RMS $\downarrow$	A.Rel $\downarrow$
BTS	0.780	8.319	0.267
AdaBins	0.750	8.686	0.195
NeWCRF	0.707	9.437	0.232
iDisc	**0.821**	**7.567**	**0.163**

Table 5. Results on DDAD proposed split.

Method	$\delta_1 \uparrow$	RMS $\downarrow$	A.Rel $\downarrow$
BTS	0.757	10.11	0.186
AdaBins	0.748	10.24	0.201
NeWCRF	0.702	10.98	0.219
iDisc	**0.809**	**8.898**	**0.163**

Quantitative results (generalization)

Table 6. Zero-shot testing SI_{log} results.

Method	SUN	Diode	Argoverse	DDAD
BTS	14.25	23.78	51.80	40.51
AdaBins	13.20	22.54	52.33	50.71
NeWCRF	11.27	18.69	46.77	44.24
iDisc	**10.91**	**18.11**	**33.35**	**29.37**

Table 7. NYU-Surface v2 official test set results.

Method	11.5° ↑	RMS ↓	Median↓
GeoNet	0.484	26.9	11.8
GeoNet++	0.502	26.7	11.2
Bae et al.	0.622	23.5	7.5
iDisc	**0.638**	**22.8**	**7.3**

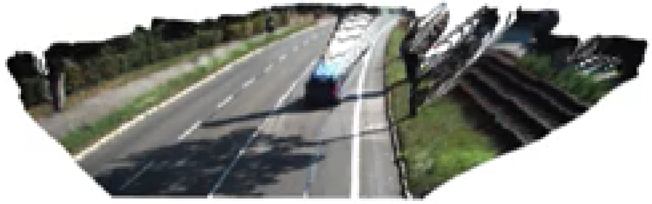
Ground Truth

AdaBins

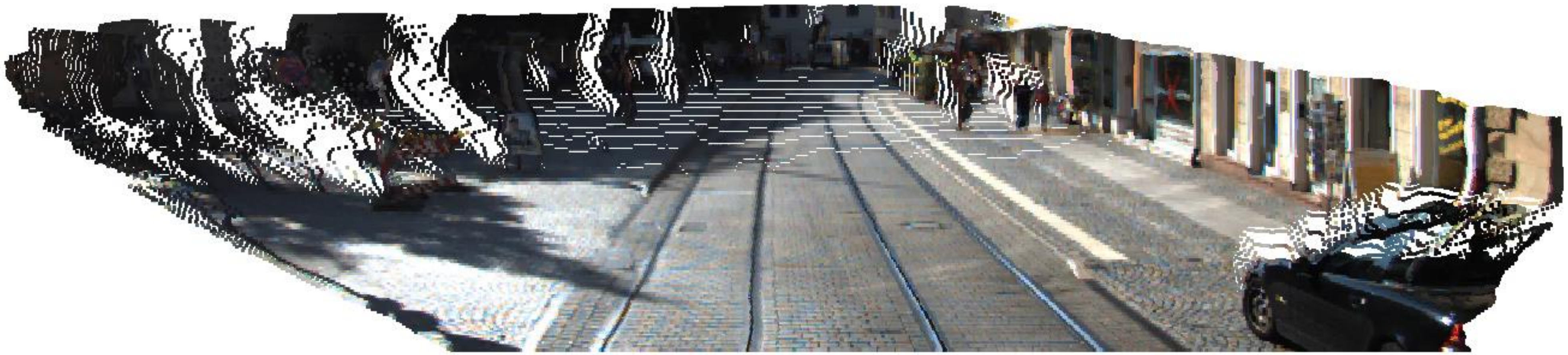
NeWCRF

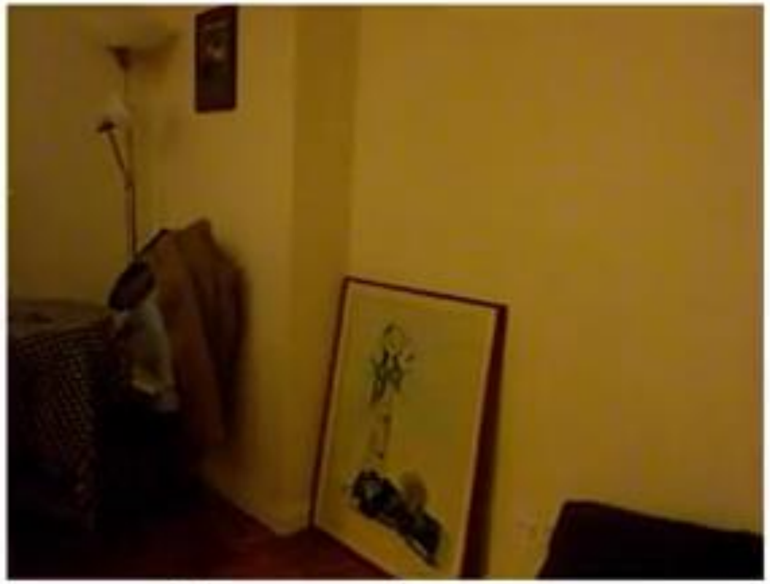
Ours

Ground Truth



Ours

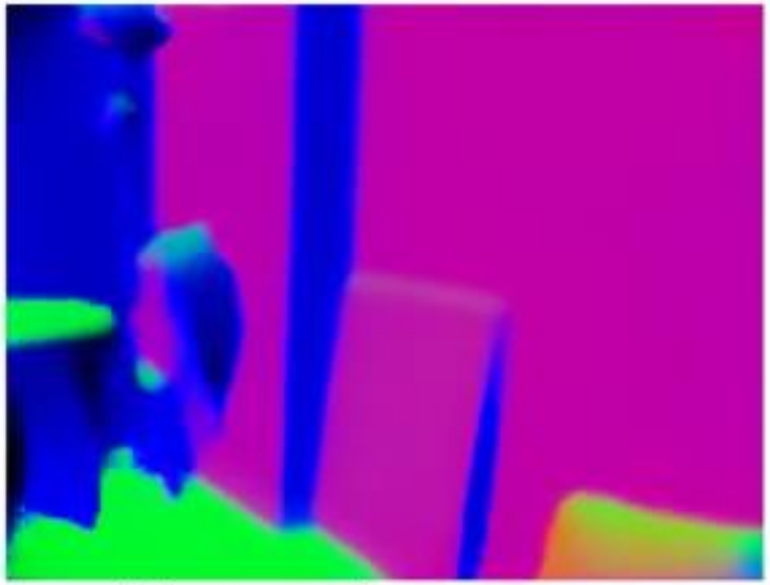




Input image

4 IDR attentions

Depth output



Normals output

Conclusion

- Despite the ill-posed problem, handcrafted biases are limiting.
- Input-dependent representations allow better generalization.
- General architecture for any dense real-valued tasks.




JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

iDisc: Internal Discretization for Monocular Depth Estimation

Luigi Piccinelli, Christos Sakaridis, Fisher Yu

Project page:
vis.xyz/pub/idisc

Code and models:
github.com/SysCV/idisc