Self-supervised learning for image-to-image translation in the small data regime

PhD thesis dissertation by **Aitor Álvarez Gila**

Directors:

Joost van de Weijer

Estibaliz Garrote

CVC, Bellaterra

2022-07-19

Universitat Autònoma de Barcelona

Computer Vision Center

MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

Self-supervised learning for image-to-image translation in the small data regime

Introduction

Overcome the fully supervised, end-to-end paradigm on large-scale annotated datasets

Leverage our prior knowledge of image formation process

Small *labeled* data domain

Chapter	Small data	Technique	Prior knowledge/Physics	Synthetic
Ch.2 (SynthBlur)	Few/Auto labels	CNN	Blurred image formation	Ground Truth
Ch.3 (RGB2HSI)	Auto labels	CNN	Color image formation	Input
Ch.4 (TES)	No labels	PP	Radiative Transfer	-
Ch.5 (MVMO), Ch.6 (ZPCVNet)	Few labels	CNN	Path tracing	(Input, Ground Truth)

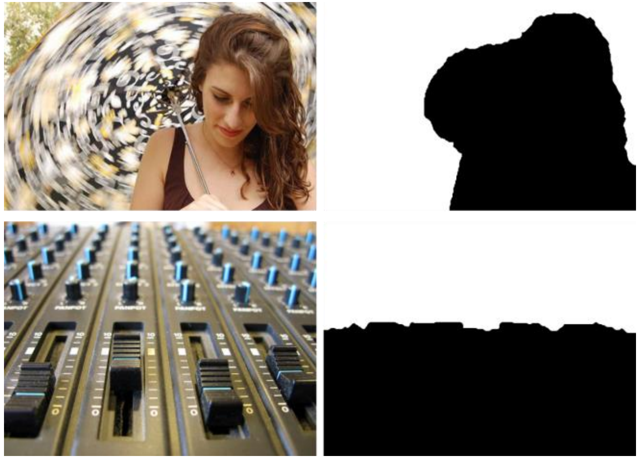
Introduction

Overcome the fully supervised, end-to-end paradigm on large-scale annotated datasets

Leverage our prior knowledge of image formation process

Small *labeled* data domain

Chapter	Small data	Technique	Prior knowledge/Physics	Synthetic
Ch.2 (SynthBlur)	Few/Auto labels	CNN	Blurred image formation	Ground Truth
Ch.3 (RGB2HSI)	Auto labels	CNN	Color image formation	Input
Ch.4 (TES)	No labels	PP	Radiative Transfer	-
Ch.5 (MVMO), Ch.6 (ZPCVNet)	Few labels	CNN	Path tracing	(Input, Ground Truth)



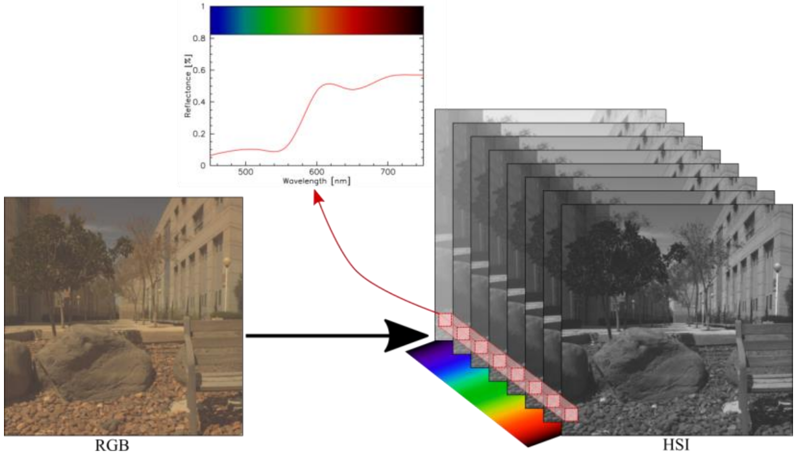
Introduction

Overcome the fully supervised, end-to-end paradigm on large-scale annotated datasets

Leverage our prior knowledge of image formation process

Small *labeled* data domain

Chapter	Small data	Technique	Prior knowledge/Physics	Synthetic
Ch.2 (SynthBlur)	Few/Auto labels	CNN	Blurred image formation	Ground Truth
Ch.3 (RGB2HSI)	Auto labels	CNN	Color image formation	Input
Ch.4 (TES)	No labels	PP	Radiative Transfer	-
Ch.5 (MVMO), Ch.6 (ZPCVNet)	Few labels	CNN	Path tracing	(Input, Ground Truth)



Introduction

Overcome the fully supervised, end-to-end paradigm on large-scale annotated datasets

Leverage our prior knowledge of image formation process

Small *labeled* data domain

Chapter	Small data	Technique	Prior knowledge/Physics	Synthetic
Ch.2 (SynthBlur)	Few/Auto labels	CNN	Blurred image formation	Ground Truth
Ch.3 (RGB2HSI)	Auto labels	CNN	Color image formation	Input
Ch.4 (TES)	No labels	PP	Radiative Transfer	-
Ch.5 (MVMO), Ch.6 (ZPCVNet)	Few labels	CNN	Path tracing	(Input, Ground Truth)

Introduction

Overcome the fully supervised, end-to-end paradigm on large-scale annotated datasets

Leverage our prior knowledge of image formation process

Small *labeled* data domain

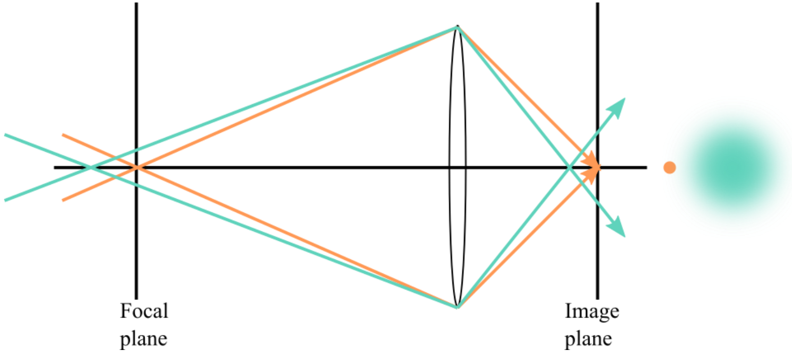
Chapter	Small data	Technique	Prior knowledge/Physics	Synthetic
Ch.2 (SynthBlur)	Few/Auto labels	CNN	Blurred image formation	Ground Truth
Ch.3 (RGB2HSI)	Auto labels	CNN	Color image formation	Input
Ch.4 (TES)	No labels	PP	Radiative Transfer	-
Ch.5 (MVMO), Ch.6 (ZPCVNet)	Few labels	CNN	Path tracing	(Input, Ground Truth)

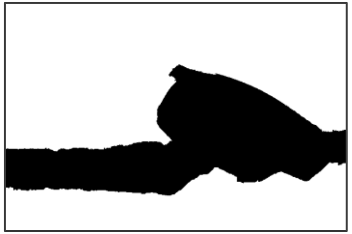
Self-supervised learning for image-to-image translation in the small data regime

Problem statement

Image blur detection for…

Defocus blur:
Wide aperture projecting scene points on a circle of confusion





Motion blur:
Object/camera movement during exposure

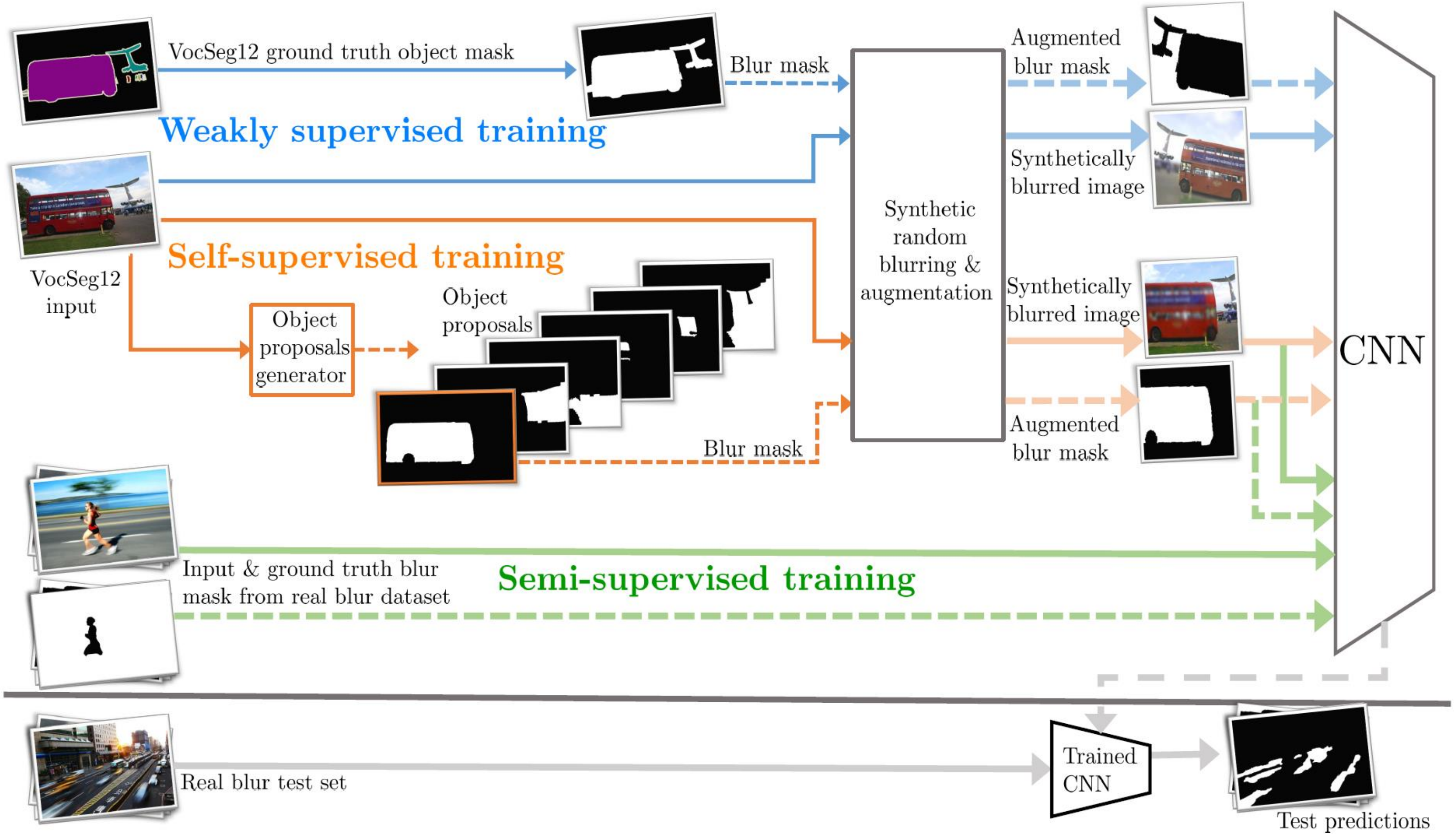
Image: phlearn.com

Fully sup progress hindered by lack of large scale datasets →

Chapter	Small data	Technique	Prior knowledge/Physics	Synthetic
Ch.2 (SynthBlur)	Few/Auto labels	CNN	Blurred image formation	Ground Truth

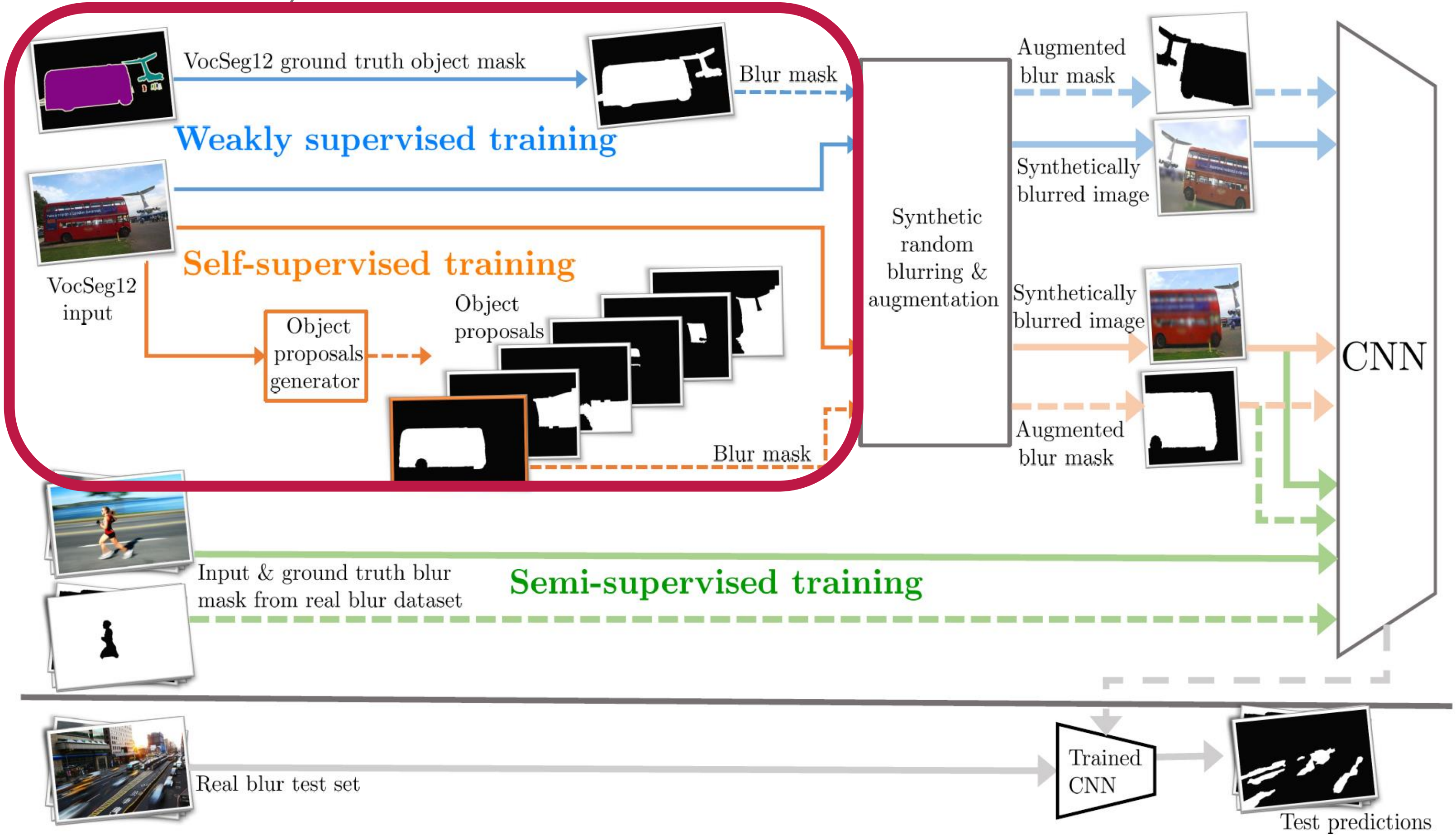
Overview

One framework, 3 instantiations

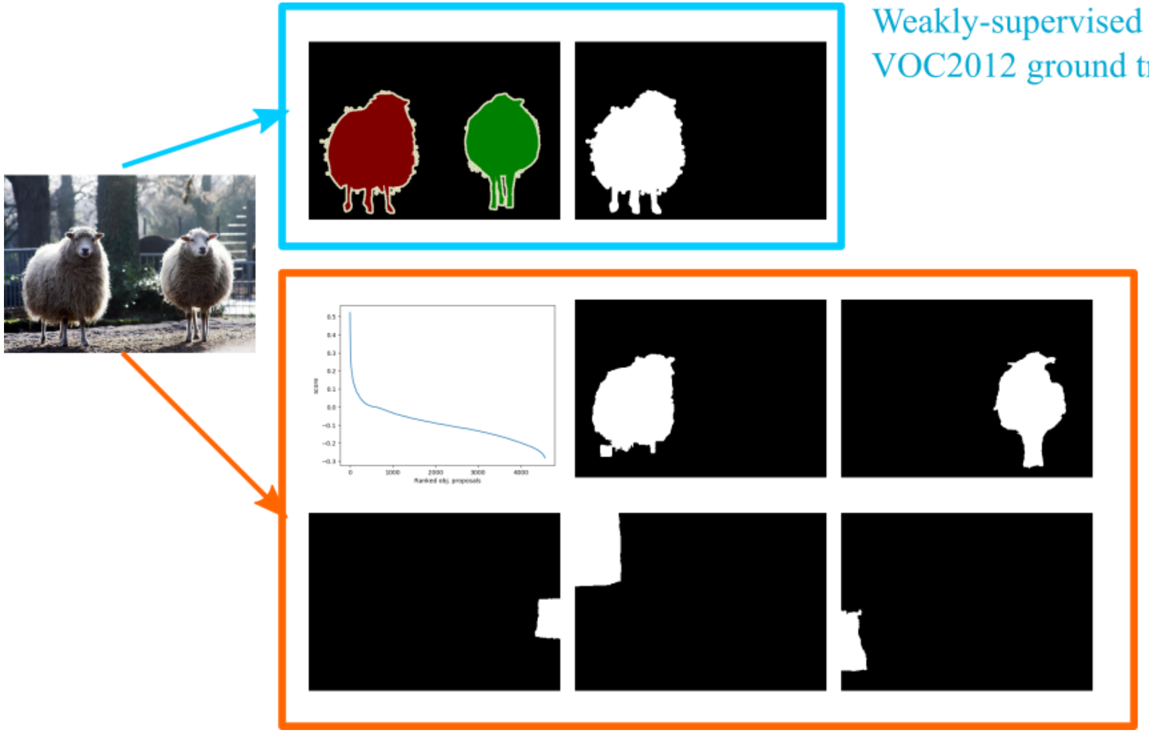


Overview

One framework, 3 instantiations



Blur mask extraction



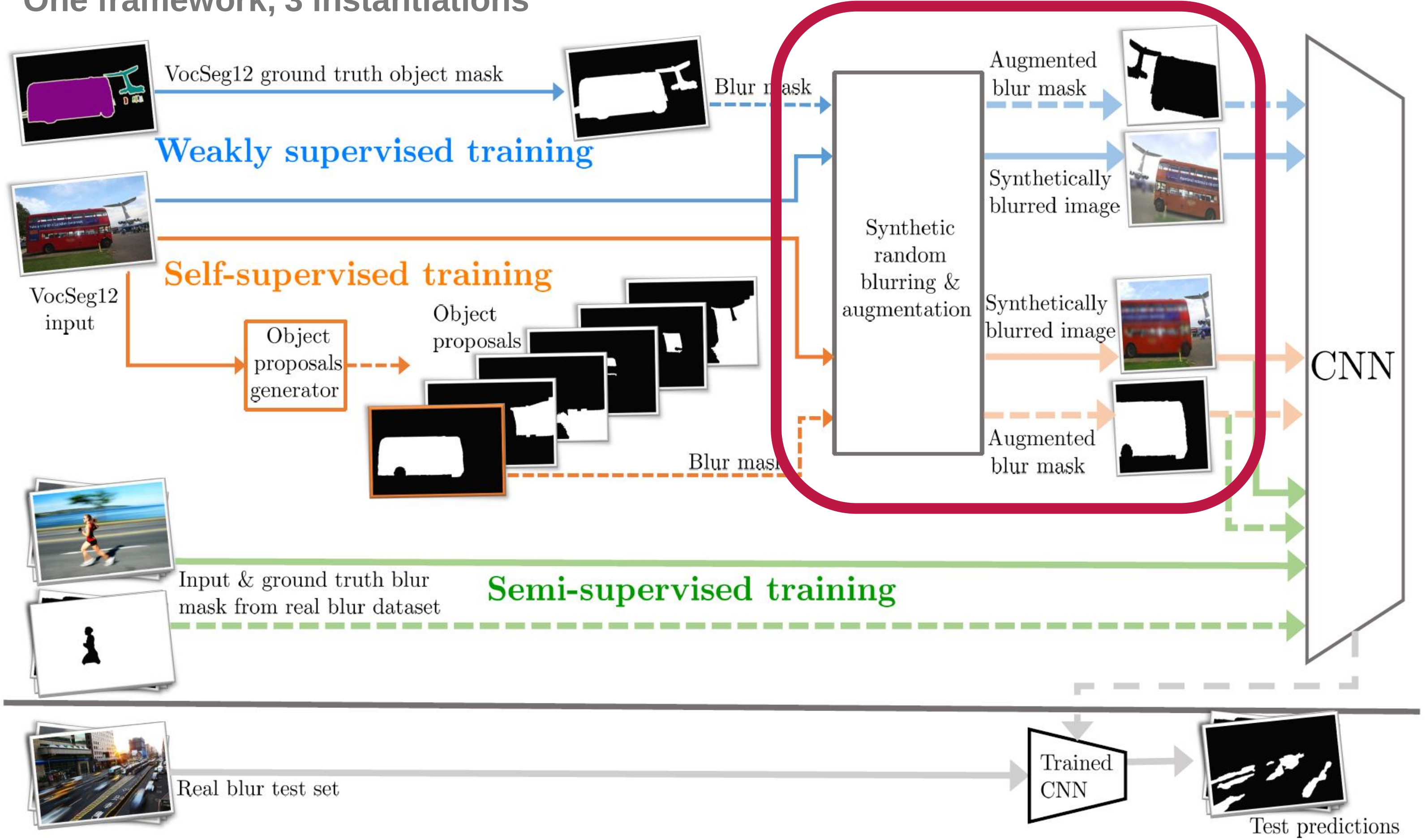
Weakly-supervised path:
VOC2012 ground truth ***semantic masks***

Self-supervised path:
Unsupervised ***object proposals*** via
Multiscale Combinatorial Grouping (MCG)
[Pont-Tuset,TPAMI 2017]

* Invert blur mask with prob p to avoid foreground bias.

Overview

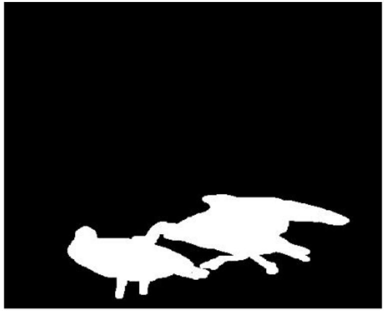
One framework, 3 instantiations



Synthetic blurring

$I(x,y)$

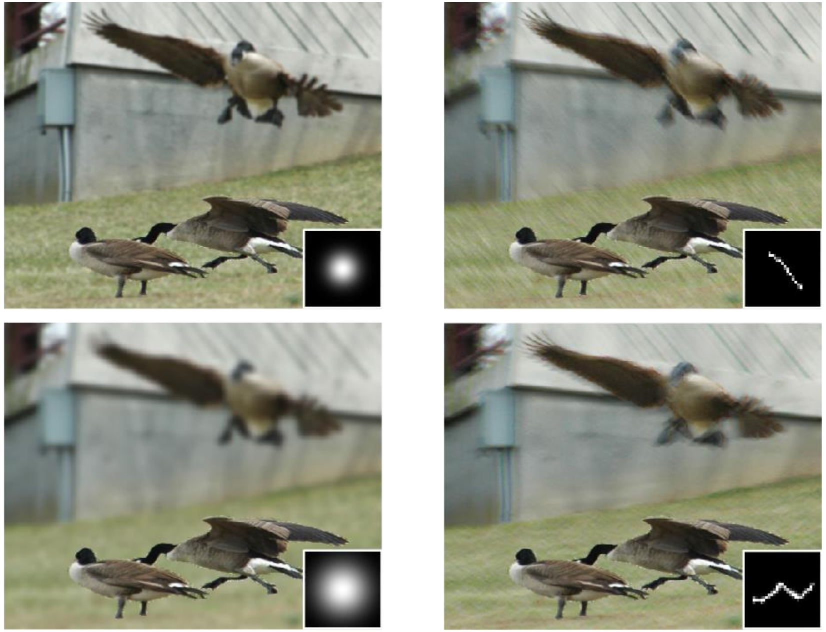
Ω^F



$I_b(x,y)$

Defocus

Motion



$$I_b(x,y) = \begin{cases} K * I(x,y), & \text{for } (x,y) \in \Omega^B, \\ I(x,y), & \text{for } (x,y) \in \Omega^F. \end{cases}$$

K – randomized blur kernel:

- Defocus: Gaussian K_σ
- Motion: non-linear motion blur kernel $K_{(m,\alpha,\mathscr{E})}$: elastic deformation over rotated line

length

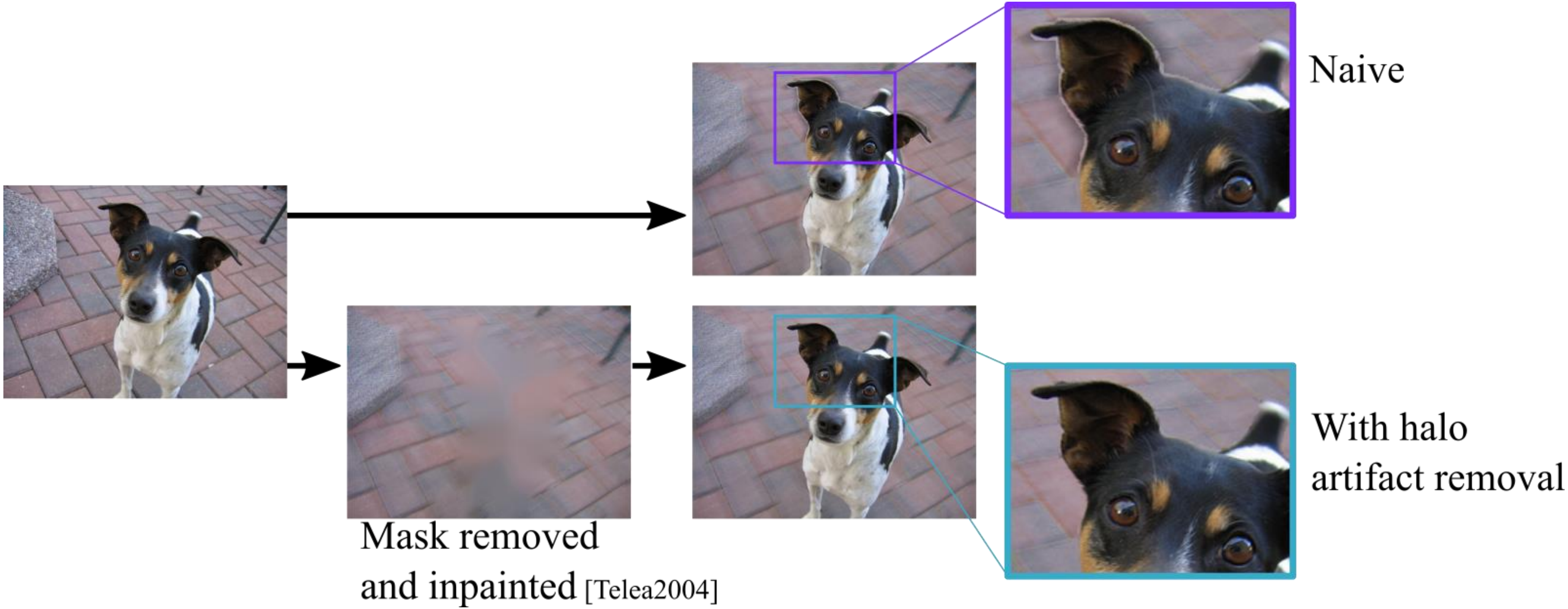
rotation angle

Elastic deformation

Synthetic blurring

Halo artifact removal through inpainting

Prevents the model from learning shortcuts for blur detection.



[Telea2004] A. Telea, "An Image Inpainting Technique Based on the Fast Marching Method," Journal of Graphics Tools, 2004.

Experiments

Evaluation on [Shi2014]'s 500 even images

Segmentation CNN: Deeplabv3[Resnet101]

- Large receptive fields (atrous convolutions)
- Multi-scale feature fusion

Best, 2nd best and 3rd best

Method	AUC Defocus	AUC Motion	AUC All	AP Defocus	AP Motion	AP All
Liu *et al.* [151]	0.722	0.714	0.720	0.792	0.683	0.760
Chakrabarti [32]	0.745	0.640	0.714	0.837	0.675	0.789
Su *et al.* [246]	0.807	0.750	0.790	0.859	0.707	0.814
Shi *et al.* [238]	0.836	0.735	0.806	0.876	0.699	0.823
LBP [284]	0.855	0.678	0.802	0.876	0.683	0.819
HiFST [81]	0.901	0.804	0.873	0.928	0.744	0.874
Ma *et al.* [159] (fully supervised ad hoc CNN)	0.947	0.861	0.922	0.966	0.784	0.912
Ours self-supervised	0.945	0.905	0.933	0.960	0.838	0.924
Ours weakly supervised (segmentation masks)	0.941	0.897	0.928	0.959	0.849	0.926
Ours semi-supervised (joint with 400 odd img.)	0.956	0.904	0.941	0.974	0.840	0.934
Fully supervised (finetuned to 400 odd img.)	0.943	0.875	0.923	0.965	0.819	0.922

[Shi2014] J. Shi et al., "Discriminative Blur Detection Features," CVPR 2014.

Experiments

Evaluation on [Shi2014]'s 500 even images

Segmentation CNN: Deeplabv3[Resnet101]

- Large receptive fields (atrous convolutions)
- Multi-scale feature fusion

Best, 2nd best and 3rd best

	AUC			AP		
Method	Defocus	Motion	All	Defocus	Motion	All
Liu *et al.* [151]	0.722	0.714	0.720	0.792	0.683	0.760
Chakrabarti [32]	0.745	0.640	0.714	0.837	0.675	0.789
Su *et al.* [246]	0.807	0.750	0.790	0.859	0.707	0.814
Shi *et al.* [238]	0.836	0.735	0.806	0.876	0.699	0.823
LBP [284]	0.855	0.678	0.802	0.876	0.683	0.819
HiFST [81]	0.901	0.804	0.873	0.928	0.744	0.874
Ma *et al.* [159] (fully supervised ad hoc CNN)	0.947	0.861	0.922	0.966	0.784	0.912
Ours self-supervised	0.945	0.905	0.933	0.960	0.838	0.924
Ours weakly supervised (segmentation masks)	0.941	0.897	0.928	0.959	0.849	0.926
Ours semi-supervised (joint with 400 odd img.)	0.956	0.904	0.941	0.974	0.840	0.934
Fully supervised (finetuned to 400 odd img.)	0.943	0.875	0.923	0.965	0.819	0.922

[Shi2014] J. Shi et al., "Discriminative Blur Detection Features," CVPR 2014.

Experiments

Evaluation on [Shi2014]'s 500 even images

Segmentation CNN: Deeplabv3[Resnet101]

- Large receptive fields (atrous convolutions)
- Multi-scale feature fusion

Best, 2nd best and 3rd best

Method	AUC			AP		
	Defocus	Motion	All	Defocus	Motion	All
Liu *et al.* [151]	0.722	0.714	0.720	0.792	0.683	0.760
Chakrabarti [32]	0.745	0.640	0.714	0.837	0.675	0.789
Su *et al.* [246]	0.807	0.750	0.790	0.859	0.707	0.814
Shi *et al.* [238]	0.836	0.735	0.806	0.876	0.699	0.823
LBP [284]	0.855	0.678	0.802	0.876	0.683	0.819
HiFST [81]	0.901	0.804	0.873	0.928	0.744	0.874
Ma *et al.* [159] (fully supervised ad hoc CNN)	0.947	0.861	0.922	0.966	0.784	0.912
Ours self-supervised	0.945	<u>0.905</u>	0.933	0.960	0.838	0.924
Ours weakly supervised (segmentation masks)	0.941	0.897	0.928	0.959	<u>0.849</u>	0.926
Ours semi-supervised (joint with 400 odd img.)	<u>0.956</u>	0.904	<u>0.941</u>	<u>0.974</u>	0.840	<u>0.934</u>
Fully supervised (finetuned to 400 odd img.)	0.943	0.875	0.923	0.965	0.819	0.922

[Shi2014] J. Shi et al., "Discriminative Blur Detection Features," CVPR 2014.

Experiments

Evaluation on [Shi2014]'s 500 even images

Segmentation CNN: Deeplabv3[Resnet101]

- Large receptive fields (atrous convolutions)
- Multi-scale feature fusion

Best, 2nd best and 3rd best

Method	AUC			AP		
	Defocus	Motion	All	Defocus	Motion	All
Liu *et al.* [151]	0.722	0.714	0.720	0.792	0.683	0.760
Chakrabarti [32]	0.745	0.640	0.714	0.837	0.675	0.789
Su *et al.* [246]	0.807	0.750	0.790	0.859	0.707	0.814
Shi *et al.* [238]	0.836	0.735	0.806	0.876	0.699	0.823
LBP [284]	0.855	0.678	0.802	0.876	0.683	0.819
HiFST [81]	0.901	0.804	0.873	0.928	0.744	0.874
Ma *et al.* [159] (fully supervised ad hoc CNN)	0.947	0.861	0.922	0.966	0.784	0.912
Ours self-supervised	0.945	0.905	0.933	0.960	0.838	0.924
Ours weakly supervised (segmentation masks)	0.941	0.897	0.928	0.959	0.849	0.926
Ours semi-supervised (joint with 400 odd img.)	0.956	0.904	0.941	0.974	0.840	0.934
Fully supervised (finetuned to 400 odd img.)	0.943	0.875	0.923	0.965	0.819	0.922

[Shi2014] J. Shi et al., "Discriminative Blur Detection Features," CVPR 2014.

Experiments

Evaluation on [Shi2014]'s 500 even images

Segmentation CNN: Deeplabv3[Resnet101]

- Large receptive fields (atrous convolutions)
- Multi-scale feature fusion

Best, 2nd best and 3rd best

Method	AUC			AP		
	Defocus	Motion	All	Defocus	Motion	All
Liu *et al.* [151]	0.722	0.714	0.720	0.792	0.683	0.760
Chakrabarti [32]	0.745	0.640	0.714	0.837	0.675	0.789
Su *et al.* [246]	0.807	0.750	0.790	0.859	0.707	0.814
Shi *et al.* [238]	0.836	0.735	0.806	0.876	0.699	0.823
LBP [284]	0.855	0.678	0.802	0.876	0.683	0.819
HiFST [81]	0.901	0.804	0.873	0.928	0.744	0.874
Ma *et al.* [159] (fully supervised ad hoc CNN)	0.947	0.861	0.922	0.966	0.784	0.912
Ours self-supervised	0.945	0.905	0.933	0.960	0.838	0.924
Ours weakly supervised (segmentation masks)	0.941	0.897	0.928	0.959	0.849	0.926
Ours semi-supervised (joint with 400 odd img.)	0.956	0.904	0.941	0.974	0.840	0.934
Fully supervised (finetuned to 400 odd img.)	0.943	0.875	0.923	0.965	0.819	0.922

[Shi2014] J. Shi et al., "Discriminative Blur Detection Features," CVPR 2014.

Experiments

Evaluation on [Shi2014]'s 500 even images

Segmentation CNN: Deeplabv3[Resnet101]

- Large receptive fields (atrous convolutions)
- Multi-scale feature fusion

Best, 2nd best and 3rd best

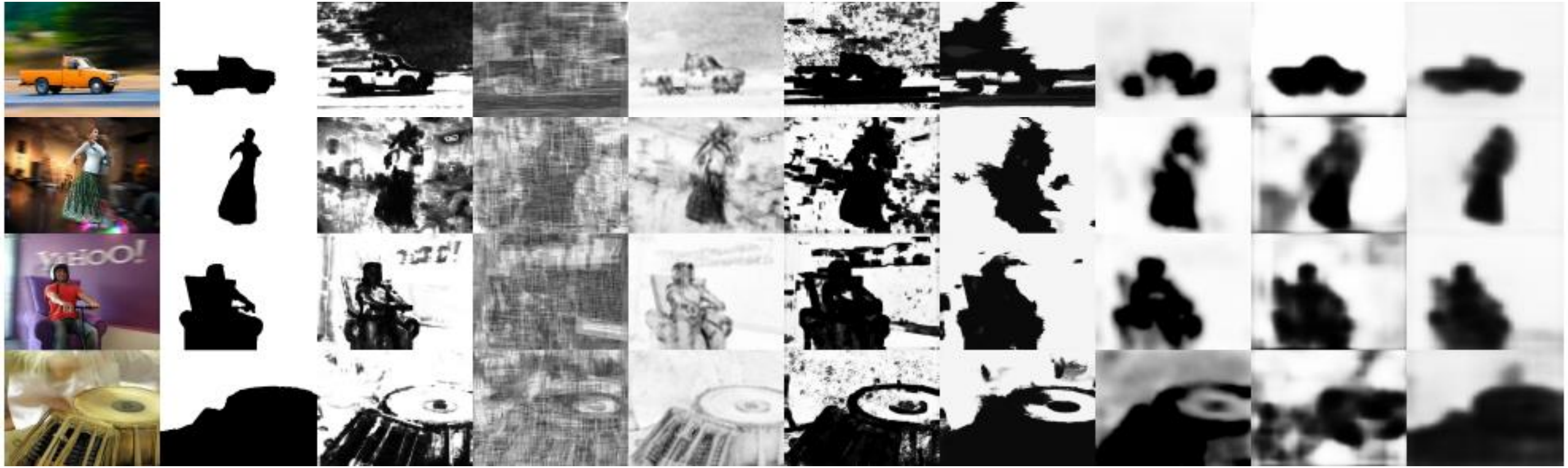
Method	AUC Defocus	AUC Motion	AUC All	AP Defocus	AP Motion	AP All
Liu *et al.* [151]	0.722	0.714	0.720	0.792	0.683	0.760
Chakrabarti [32]	0.745	0.640	0.714	0.837	0.675	0.789
Su *et al.* [246]	0.807	0.750	0.790	0.859	0.707	0.814
Shi *et al.* [238]	0.836	0.735	0.806	0.876	0.699	0.823
LBP [284]	0.855	0.678	0.802	0.876	0.683	0.819
HiFST [81]	0.901	0.804	0.873	0.928	0.744	0.874
Ma *et al.* [159] (fully supervised ad hoc CNN)	0.947	0.861	0.922	0.966	0.784	0.912
Ours self-supervised	0.945	0.905	0.933	0.960	0.838	0.924
Ours weakly supervised (segmentation masks)	0.941	0.897	0.928	0.959	0.849	0.926
Ours semi-supervised (joint with 400 odd img.)	0.956	0.904	0.941	0.974	0.840	0.934
Fully supervised (finetuned to 400 odd img.)	0.943	0.875	0.923	0.965	0.819	0.922

- Our methods beat non-deep and fully supervised deep ad-hoc CNNs (especially motion blur)
- It's not the architecture
- Self and weak supervision ~on pair

[Shi2014] J. Shi et al., "Discriminative Blur Detection Features," CVPR 2014.

Experiments

Evaluation on [Shi2014]'s 500 even images

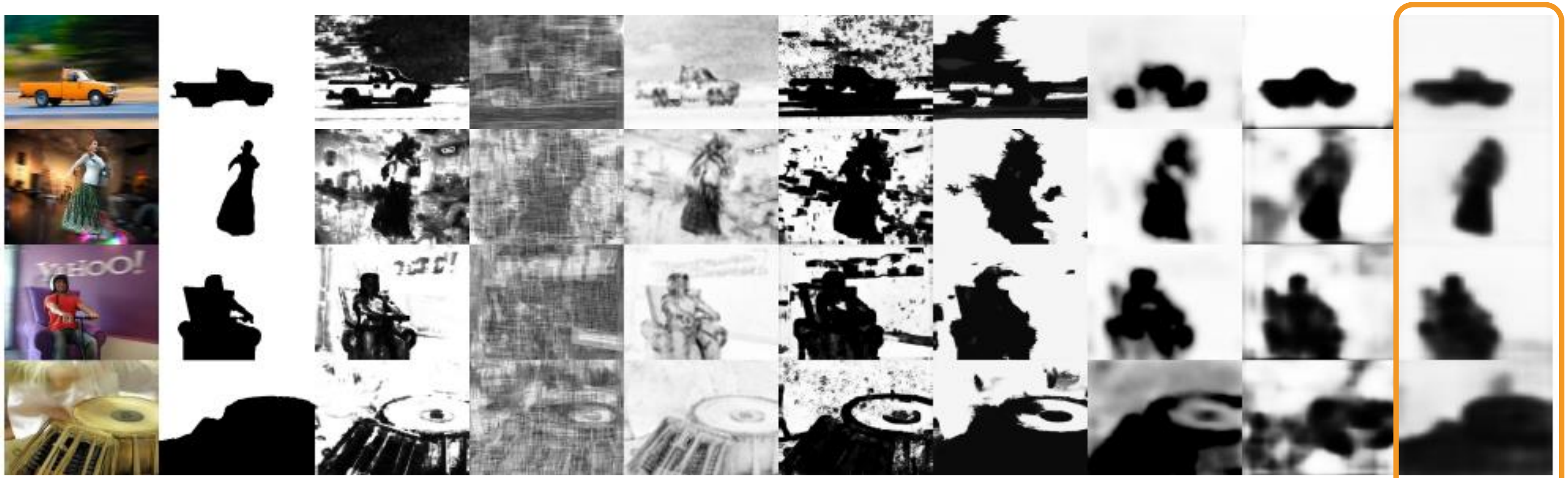


Input | Ground truth | Liu [151] | Chakraba. [32] | Su [246] | Shi [238] | LBP [284] | HiFST [81] | Ma [159] | Ours self-sup.

[Shi2014] J. Shi et al., "Discriminative Blur Detection Features," CVPR 2014.

Experiments

Evaluation on [Shi2014]'s 500 even images

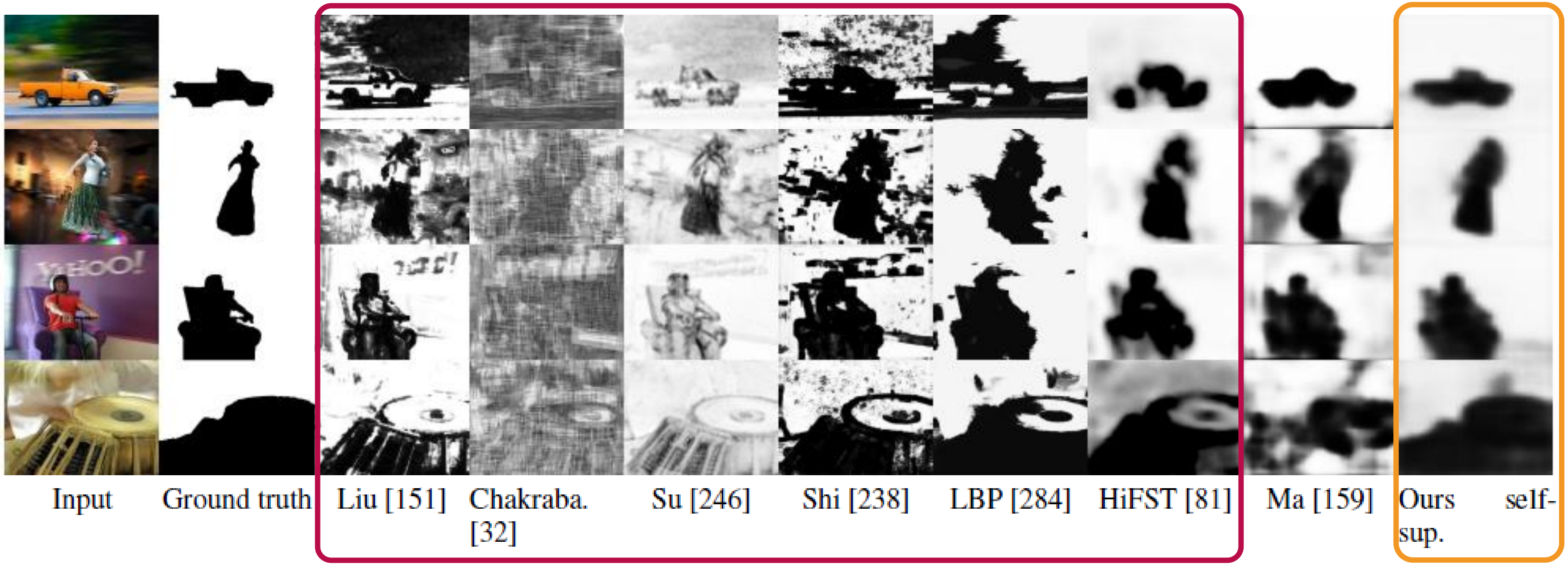


Input | Ground truth | Liu [151] | Chakraba. [32] | Su [246] | Shi [238] | LBP [284] | HiFST [81] | Ma [159] | Ours self-sup.

[Shi2014] J. Shi et al., "Discriminative Blur Detection Features," CVPR 2014.

Experiments

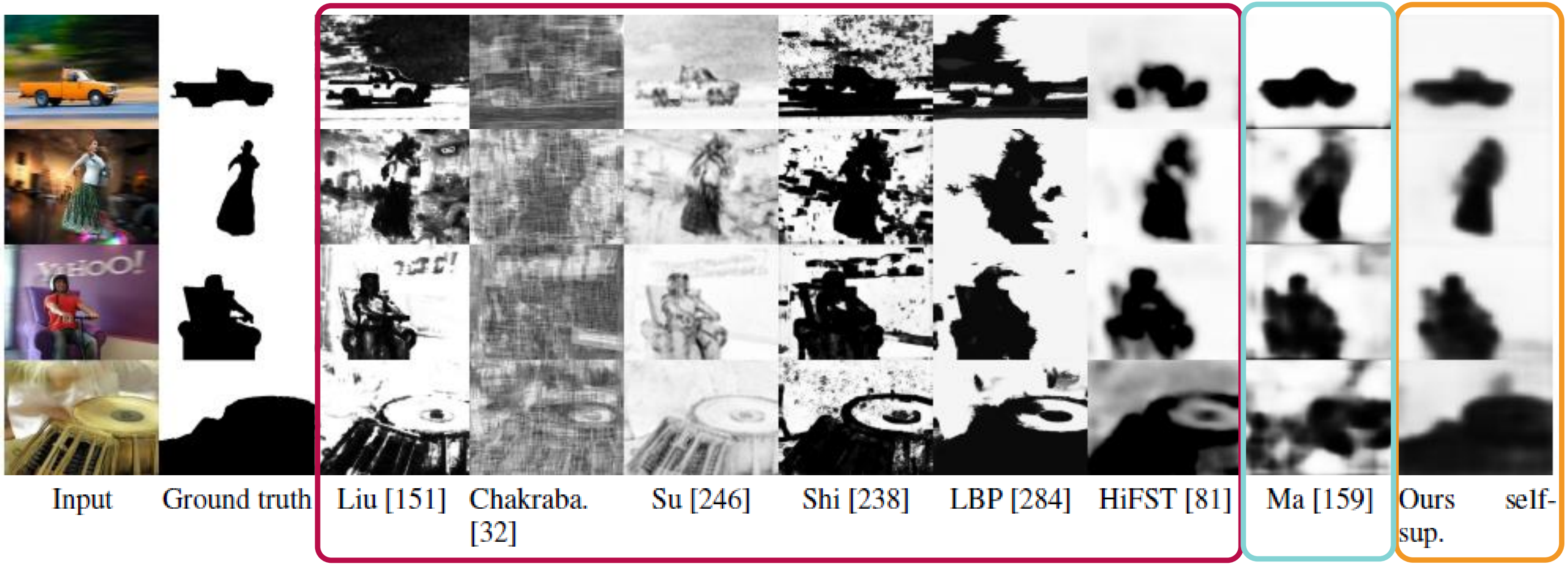
Evaluation on [Shi2014]'s 500 even images



[Shi2014] J. Shi et al., "Discriminative Blur Detection Features," CVPR 2014.

Experiments

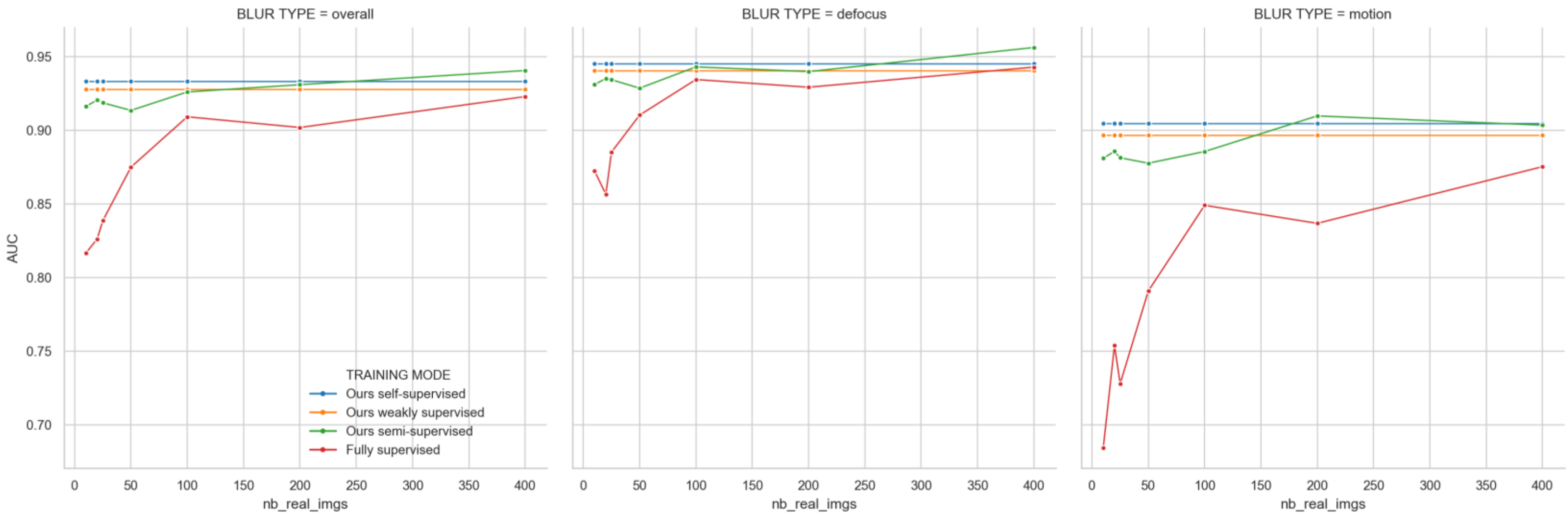
Evaluation on [Shi2014]'s 500 even images



[Shi2014] J. Shi et al., "Discriminative Blur Detection Features," CVPR 2014.

Experiments

Evaluation on [Shi2014]'s 500 even images. Semi-supervised setup (joint training)



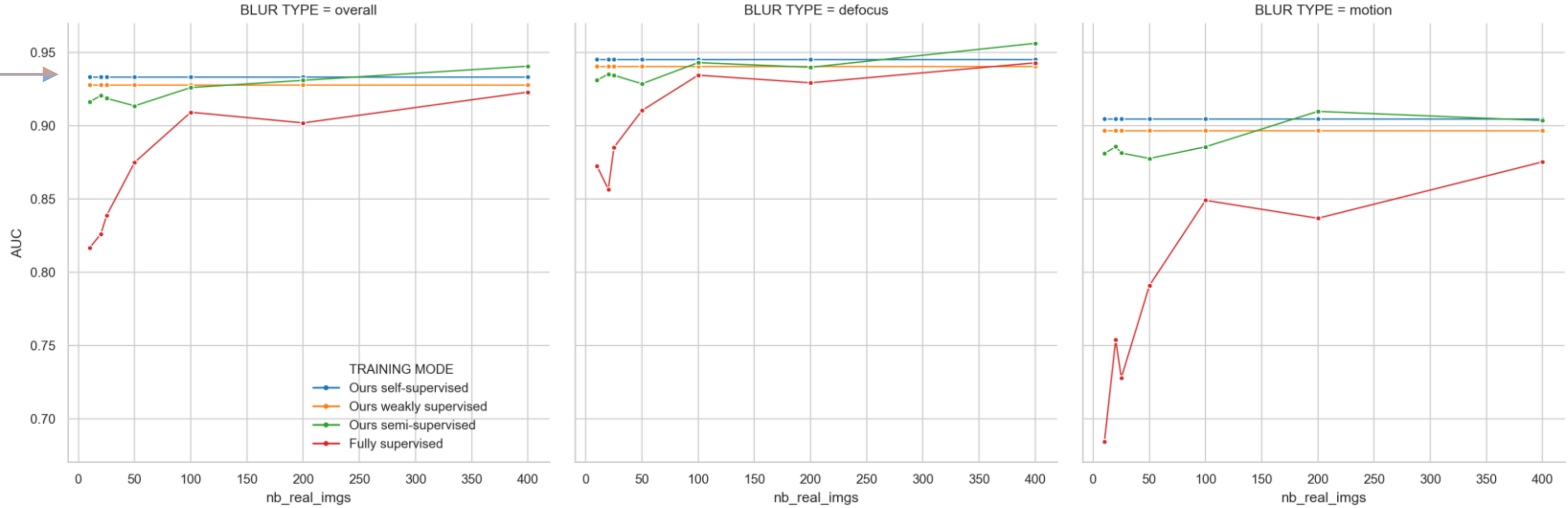
2-SynthBlur

Experiments

Evaluation on [Shi2014]’s 500 even images. Semi-supervised setup (joint training)

Do not depend on real blur masks.

Useful with few blurred images



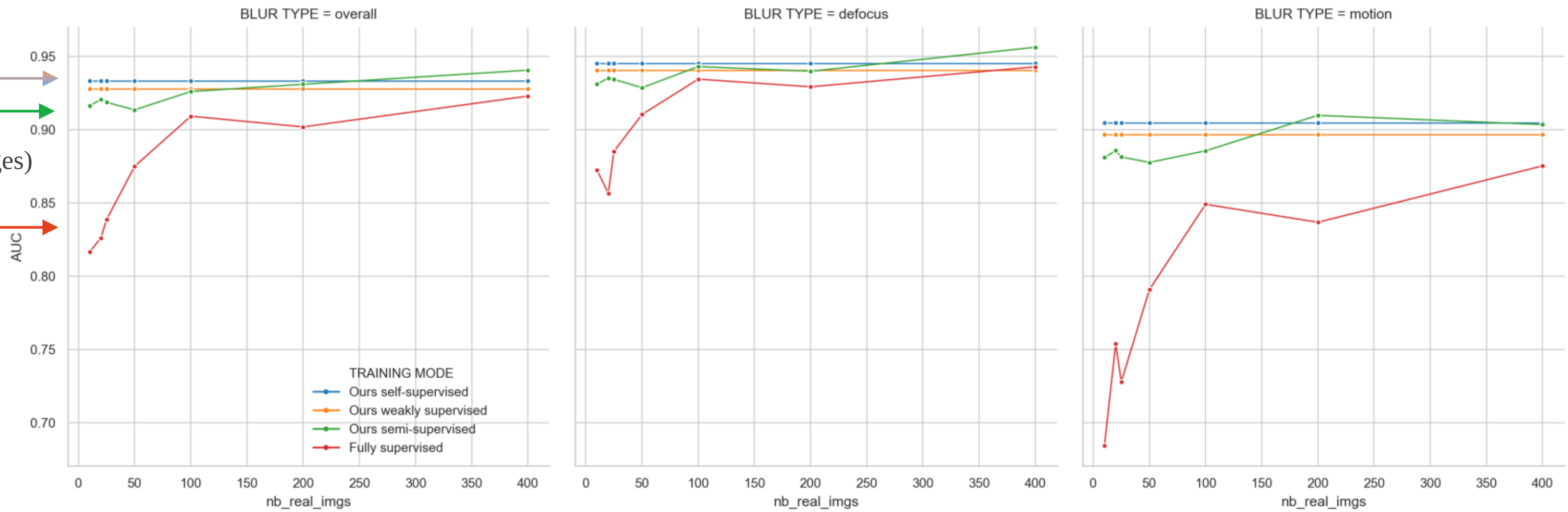
Experiments

Evaluation on [Shi2014]'s 500 even images. Semi-supervised setup (joint training)

Do not depend on real blur masks.

Useful with few blurred images

Joint > fully supervised
(Especially with few images)



Experiments

Cross-dataset generalization

Method	AUC	AP
Zhao *et al*. [294]	0.913	0.946
Ma *et al*. [159]	0.923	0.956
Ours self-supervised	**0.950**	**0.976**
Ours weakly supervised	0.915	0.953
Fully supervised	0.904	0.952

Robust generalization

[Shi2014]→[Zhao2018] direct transfer.

Defocus blur-only dataset

[Zhao2018] W. Zhao, "Defocus Blur Detection via Multi-Stream Bottom-Top-Bottom Fully Convolutional Network," CVPR 2018.

Takeaways

- Framework for defocus and motion blur segmentation from procedural synthetic local (semantically coherent) blurring. Instantiations:
  - Self-supervised
  - Weakly-supervised
  - Semi-supervised

No need for blur segmentation labels

- Good generalization
- Useful for few-data domains , e.g. medical, text, multi-spectral

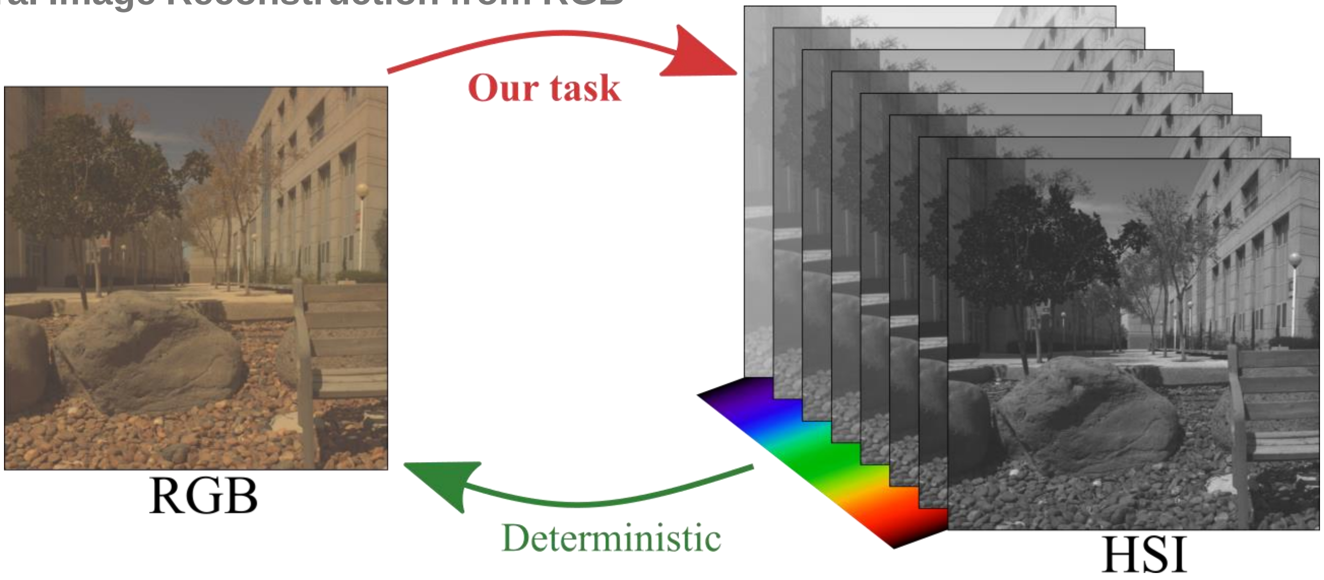
Publication **A. Alvarez-Gila**, A. Galdran, E. Garrote, and J. van de Weijer, “*Self-supervised blur detection from synthetically blurred scenes,*” Image and Vision Computing, 2019.

https://github.com/aitorshuffle/synthblur

Self-supervised learning for image-to-image translation in the small data regime

The task

Spectral Image Reconstruction from RGB



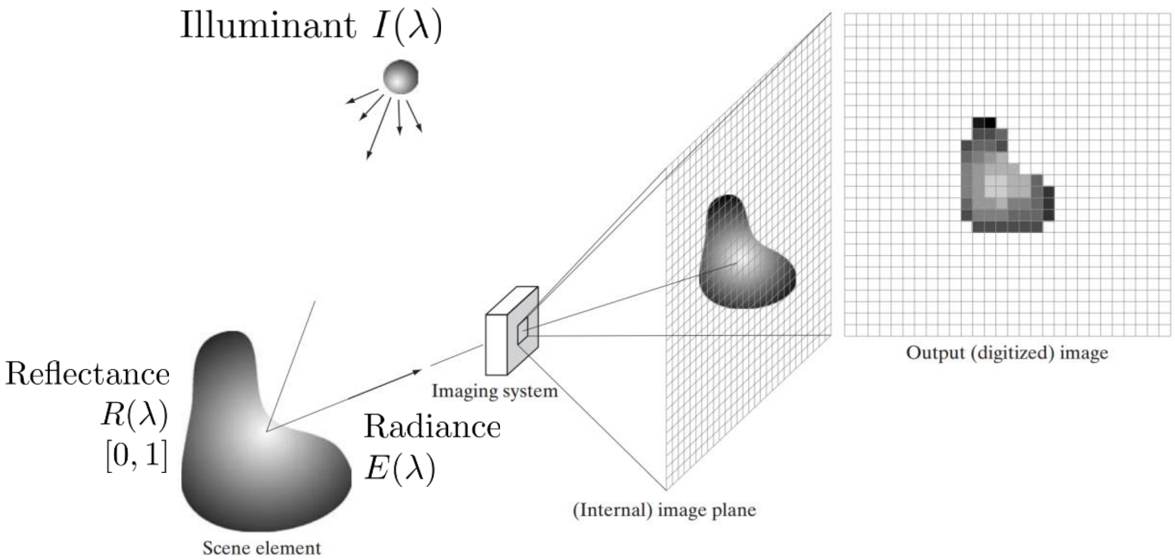
Learn a image-to-image mapping s.t., for each pixel: $\mathbb{R}^3 \rightarrow \mathbb{R}^c$ with $c >> 3$

In particular: $c = 31 \quad (400 : 10 : 700nm)$

Heavily underconstrained (e.g. metamers), non-linear problem

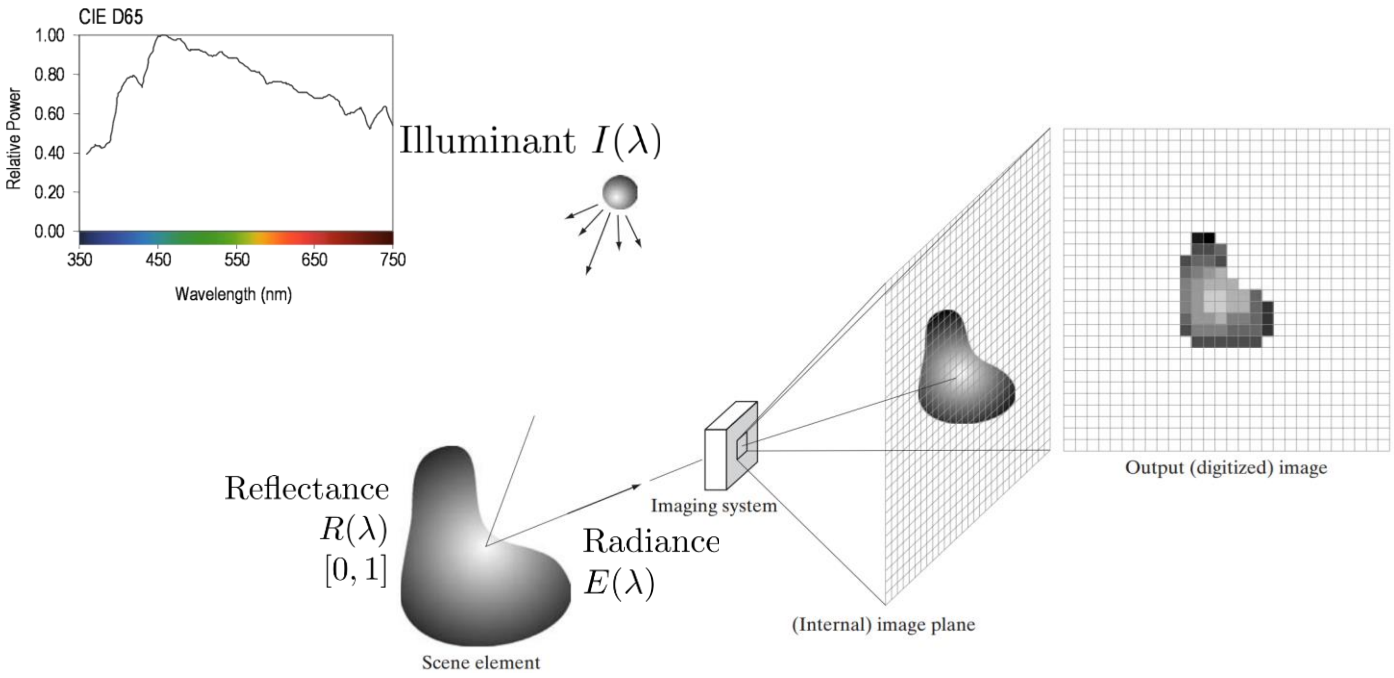
Chapter	Small data	Technique	Prior knowledge/Physics	Synthetic
Ch.3 (RGB2HSI)	Auto labels	CNN	Color image formation	Input

Image formation



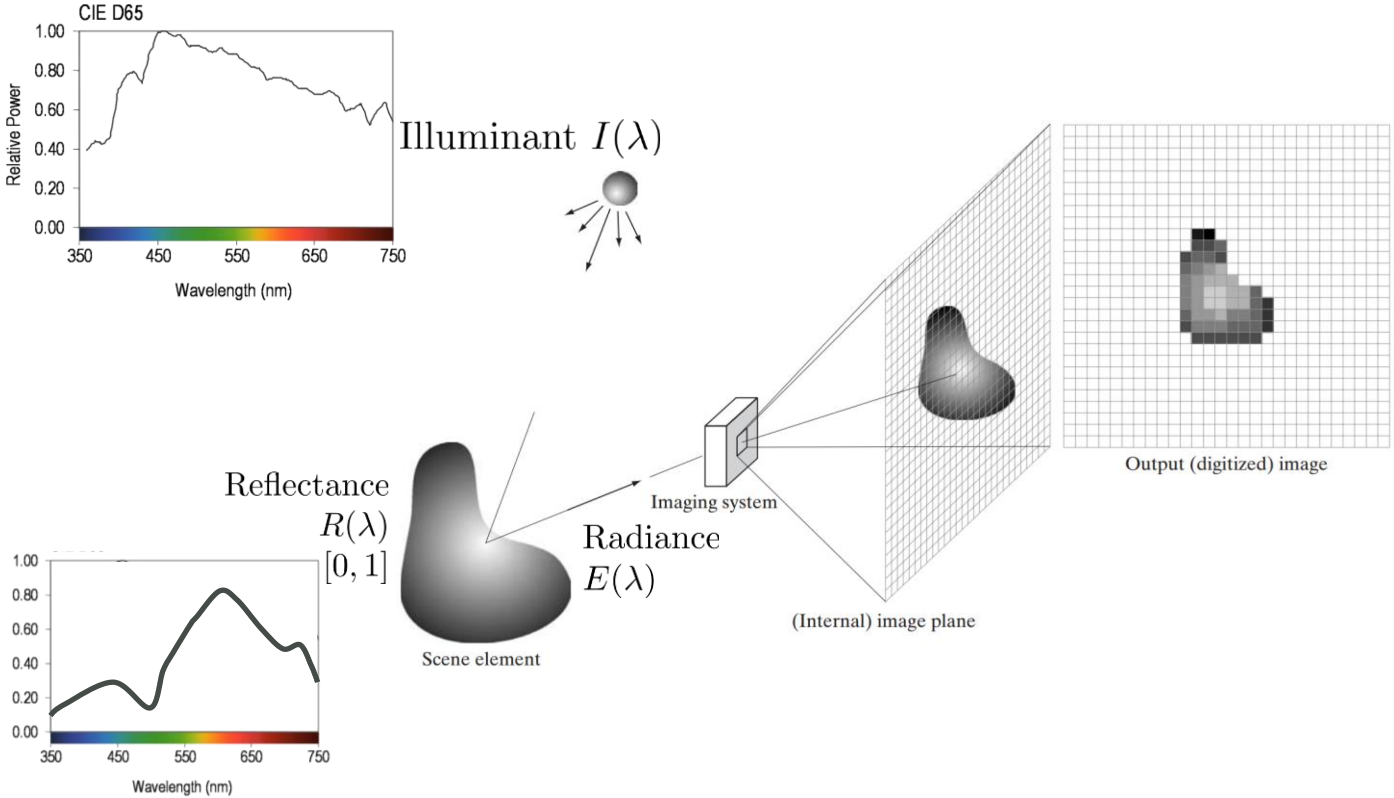
Images: R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed., Pearson, 2007.
M. S. Brown, "Understanding the In-Camera Image Processing Pipeline for Computer Vision," CVPR - Tutorial, 2016.

Image formation



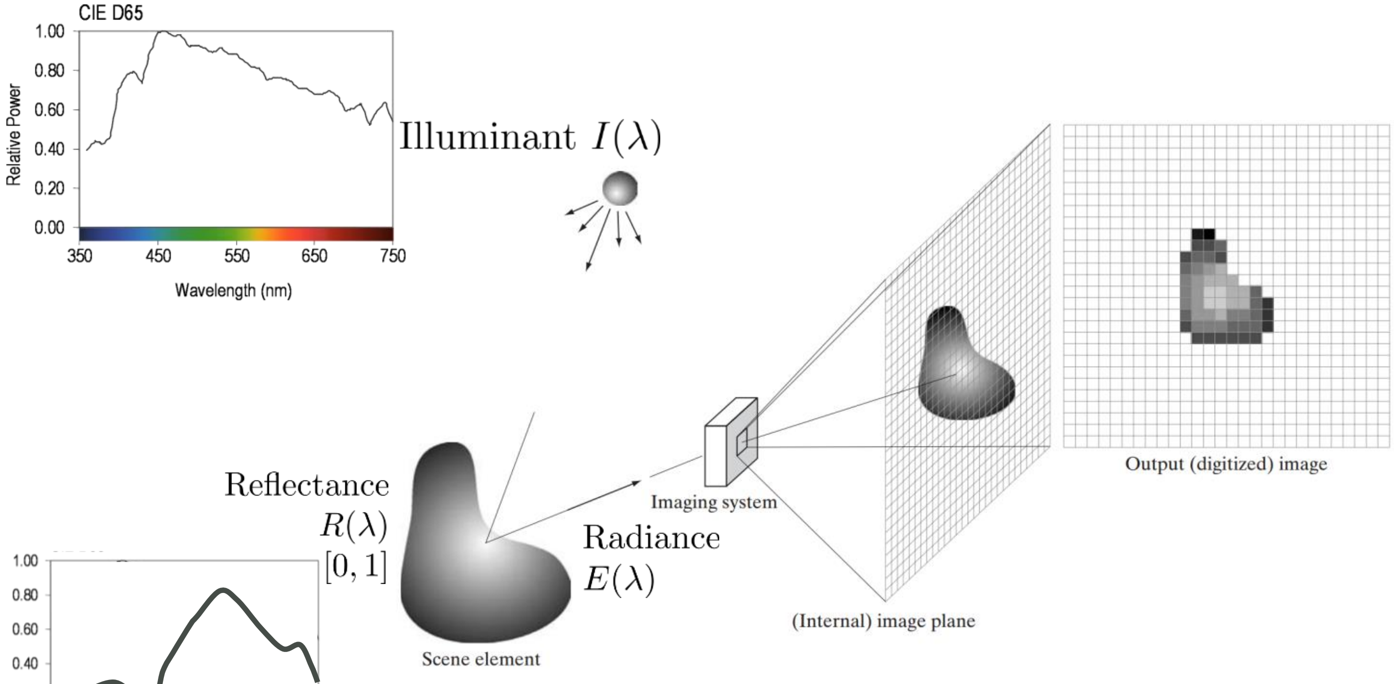
Images: R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed., Pearson, 2007.
M. S. Brown, "Understanding the In-Camera Image Processing Pipeline for Computer Vision," CVPR - Tutorial, 2016.

Image formation



Images: R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed., Pearson, 2007.
M. S. Brown, "Understanding the In-Camera Image Processing Pipeline for Computer Vision," CVPR - Tutorial, 2016.

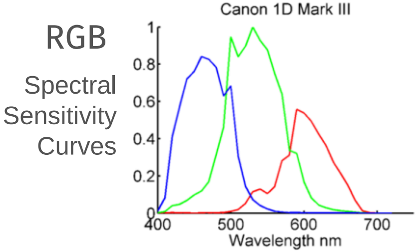
Image formation

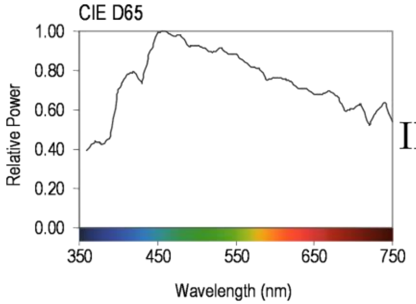


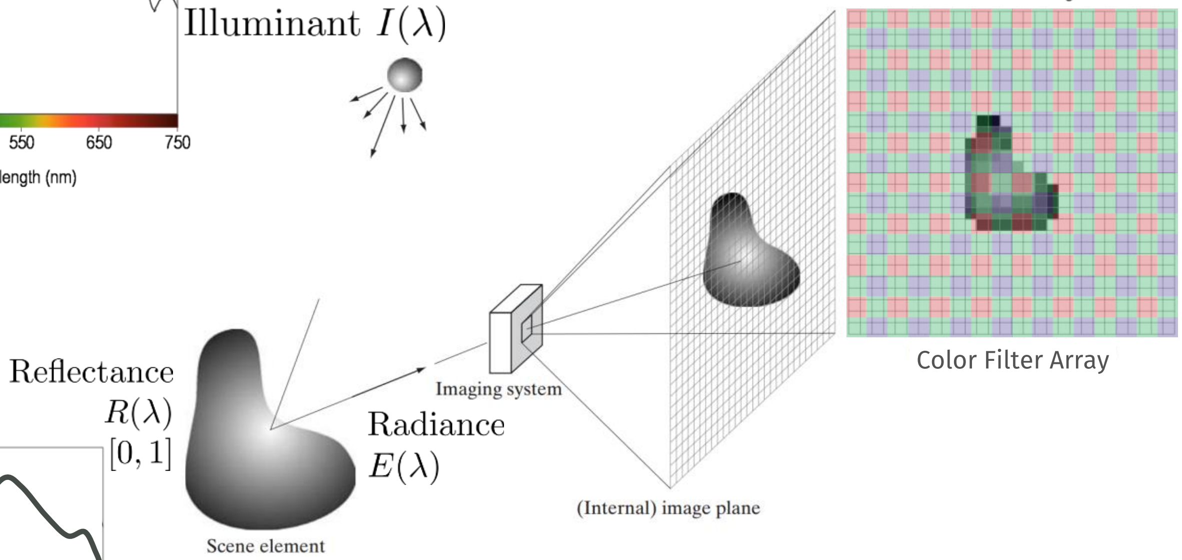
$$E(\lambda) = I(\lambda) \times R(\lambda)$$

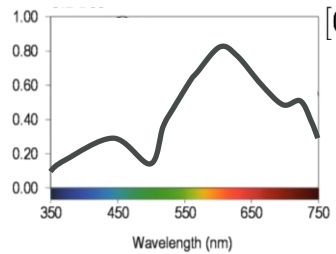
Images: R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed., Pearson, 2007.
M. S. Brown, “Understanding the In-Camera Image Processing Pipeline for Computer Vision,” CVPR - Tutorial, 2016.

Image formation





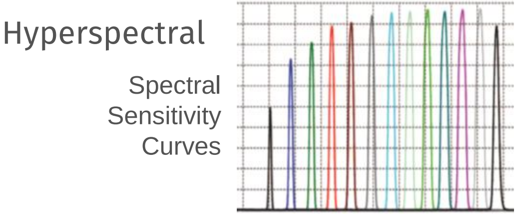


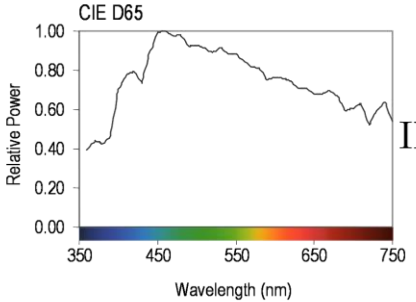


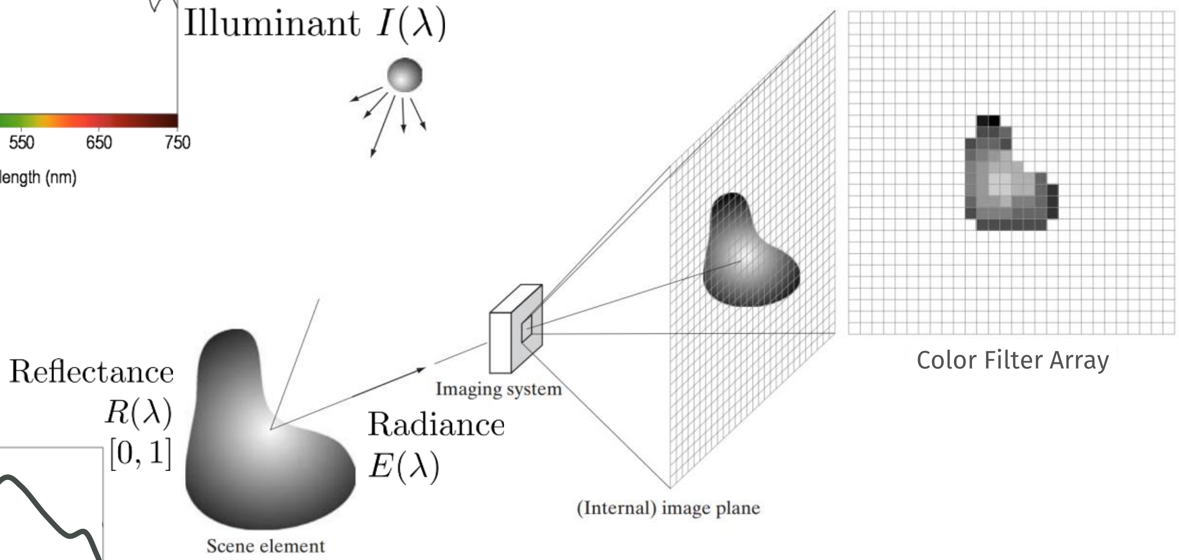
$$E(\lambda) = I(\lambda) \times R(\lambda)$$

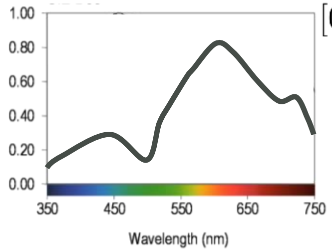
Images: R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed., Pearson, 2007.
M. S. Brown, "Understanding the In-Camera Image Processing Pipeline for Computer Vision," CVPR - Tutorial, 2016.

Image formation









$$E(\lambda) = I(\lambda) \times R(\lambda)$$

Images: R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed., Pearson, 2007.
M. S. Brown, "Understanding the In-Camera Image Processing Pipeline for Computer Vision," CVPR - Tutorial, 2016.

Related work

Spectral reconstruction

- From RGB + extra help
  - Low-res HS image:

[Cao2011] X. Cao et al., "High resolution multispectral video capture with a hybrid camera system," CVPR 2011

  - Multiplexed light:

[Park2007] J. I. Park et al., "Multispectral Imaging Using Multiplexed Illumination," ICCV 2007

[Parmar2008] M. Parmar et al., "Spatio-spectral reconstruction of the multispectral datacube using sparse recovery," ICIP 2008

[Goel2015] M. Goel et al., "HyperCam: Hyperspectral Imaging for Ubiquitous Computing Applications," IJCPUC 2015

- No spatial info considered:

[Nguyen2014] R. M. H. Nguyen et al., "Training-Based Spectral Reconstruction from a Single RGB Image," ECCV 2014

[Arad2016] B. Arad et al., "Sparse Recovery of Hyperspectral Signal from Natural RGB Images," *ECCV* 2016

- Leveraging spatial context:

[Robles-Kelly2015] A. Robles-Kelly, "Single Image Spectral Reconstruction for Multimedia Applications," ACM Multimedia, 2015.

[Galliani2017] S. Galliani et al., "Learned Spectral Super-Resolution," arXiv:1703.09470 [cs], 2017.

Related work

Spectral reconstruction

- From RGB + extra help
  - Low-res HS image:

[Cao2011] X. Cao et al., "High resolution multispectral video capture with a hybrid camera system," CVPR 2011

  - Multiplexed light:

[Park2007] J. I. Park et al., "Multispectral Imaging Using Multiplexed Illumination," ICCV 2007

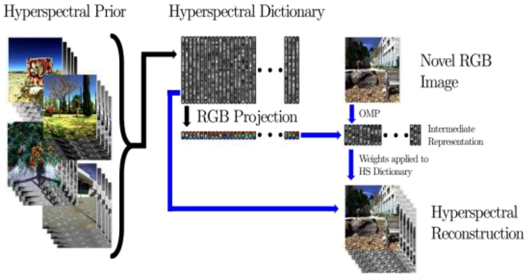
[Parmar2008] M. Parmar et al., "Spatio-spectral reconstruction of the multispectral datacube using sparse recovery," ICIP 2008

[Goel2015] M. Goel et al., "HyperCam: Hyperspectral Imaging for Ubiquitous Computing Applications," IJCPUC 2015

- No spatial info considered:

[Nguyen2014] R. M. H. Nguyen et al., "Training-Based Spectral Reconstruction from a Single RGB Image," ECCV 2014

[Arad2016] B. Arad et al., "Sparse Recovery of Hyperspectral Signal from Natural RGB Images," *ECCV* 2016



- Leveraging spatial context:

[Robles-Kelly2015] A. Robles-Kelly, "Single Image Spectral Reconstruction for Multimedia Applications," ACM Multimedia, 2015.

[Galliani2017] S. Galliani et al., "Learned Spectral Super-Resolution," arXiv:1703.09470 [cs], 2017.

Related work

Spectral reconstruction

- From RGB + extra help
  - Low-res HS image:

[Cao2011] X. Cao et al., "High resolution multispectral video capture with a hybrid camera system," CVPR 2011

  - Multiplexed light:

[Park2007] J. I. Park et al., "Multispectral Imaging Using Multiplexed Illumination," ICCV 2007

[Parmar2008] M. Parmar et al., "Spatio-spectral reconstruction of the multispectral datacube using sparse recovery," ICIP 2008

[Goel2015] M. Goel et al., "HyperCam: Hyperspectral Imaging for Ubiquitous Computing Applications," IJCPUC 2015

- No spatial info considered:

[Nguyen2014] R. M. H. Nguyen et al., "Training-Based Spectral Reconstruction from a Single RGB Image," ECCV 2014

[Arad2016] B. Arad et al., "Sparse Recovery of Hyperspectral Signal from Natural RGB Images," *ECCV* 2016

- Leveraging spatial context:

[Robles-Kelly2015] A. Robles-Kelly, "Single Image Spectral Reconstruction for Multimedia Applications," ACM Multimedia, 2015.

[Galliani2017] S. Galliani et al., "Learned Spectral Super-Resolution," arXiv:1703.09470 [cs], 2017.

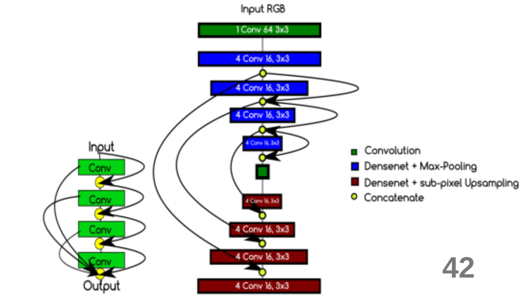


Figure 3. Depiction of a single Densenet block

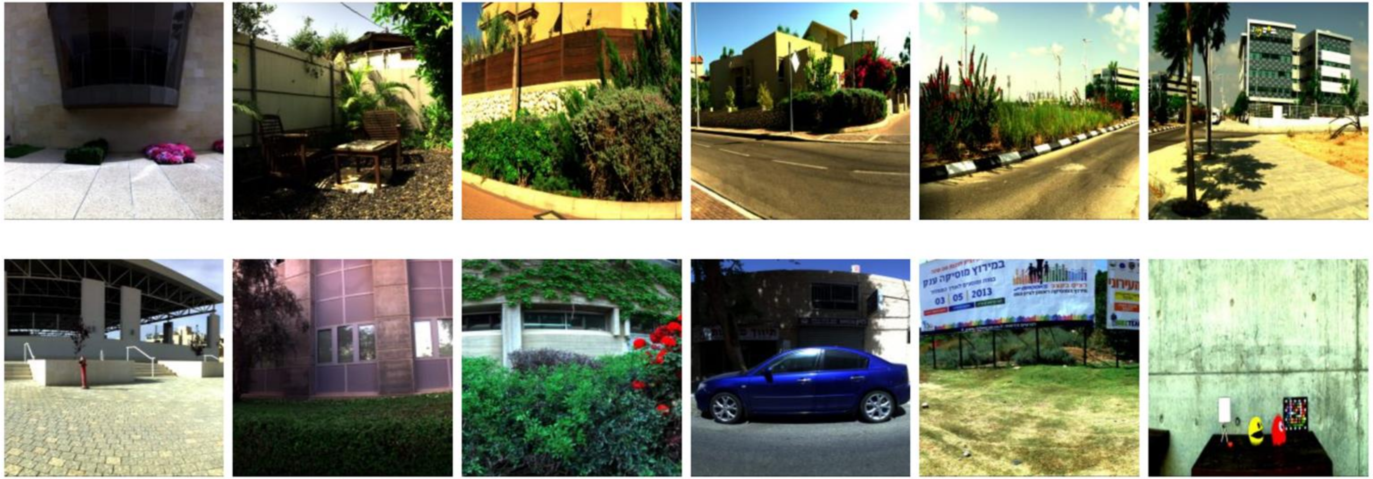
The ICVL dataset
[Arad2016]

~200 images

Camera: Specim PS Kappa DX4 + rotary

Raw HSI: 1392×1300 pixels, 519 spectral bands [400-1,000nm] with $\Delta\lambda \cong 1{,}25nm$

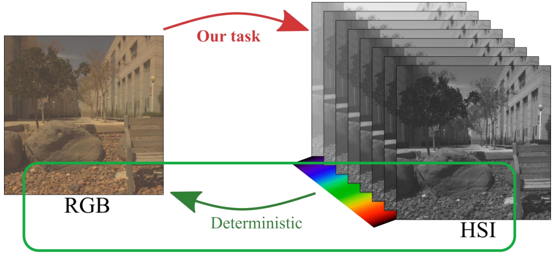
Downsampled: 31 spectral channels [400nm-700nm] with $\Delta\lambda \cong 10nm$

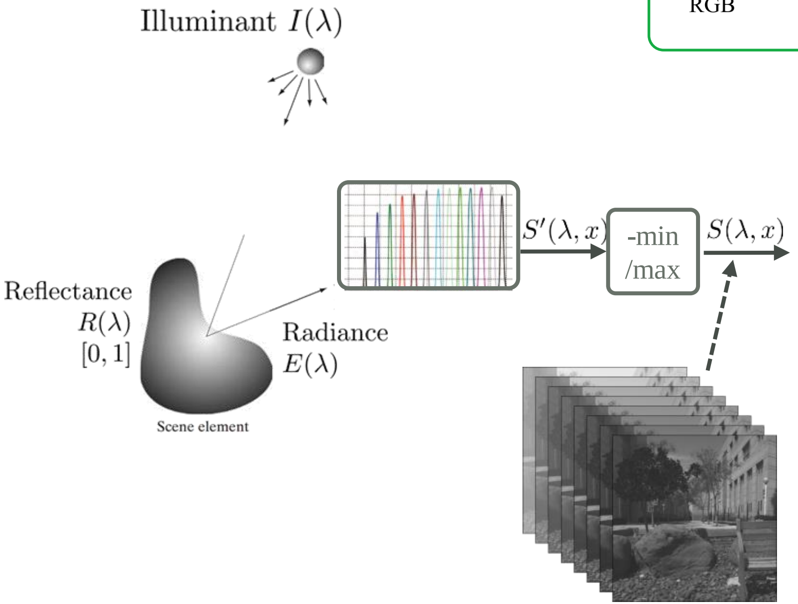


[Arad2016] B. Arad et al., "Sparse Recovery of Hyperspectral Signal from Natural RGB Images," *ECCV* 2016

Dataset preparation

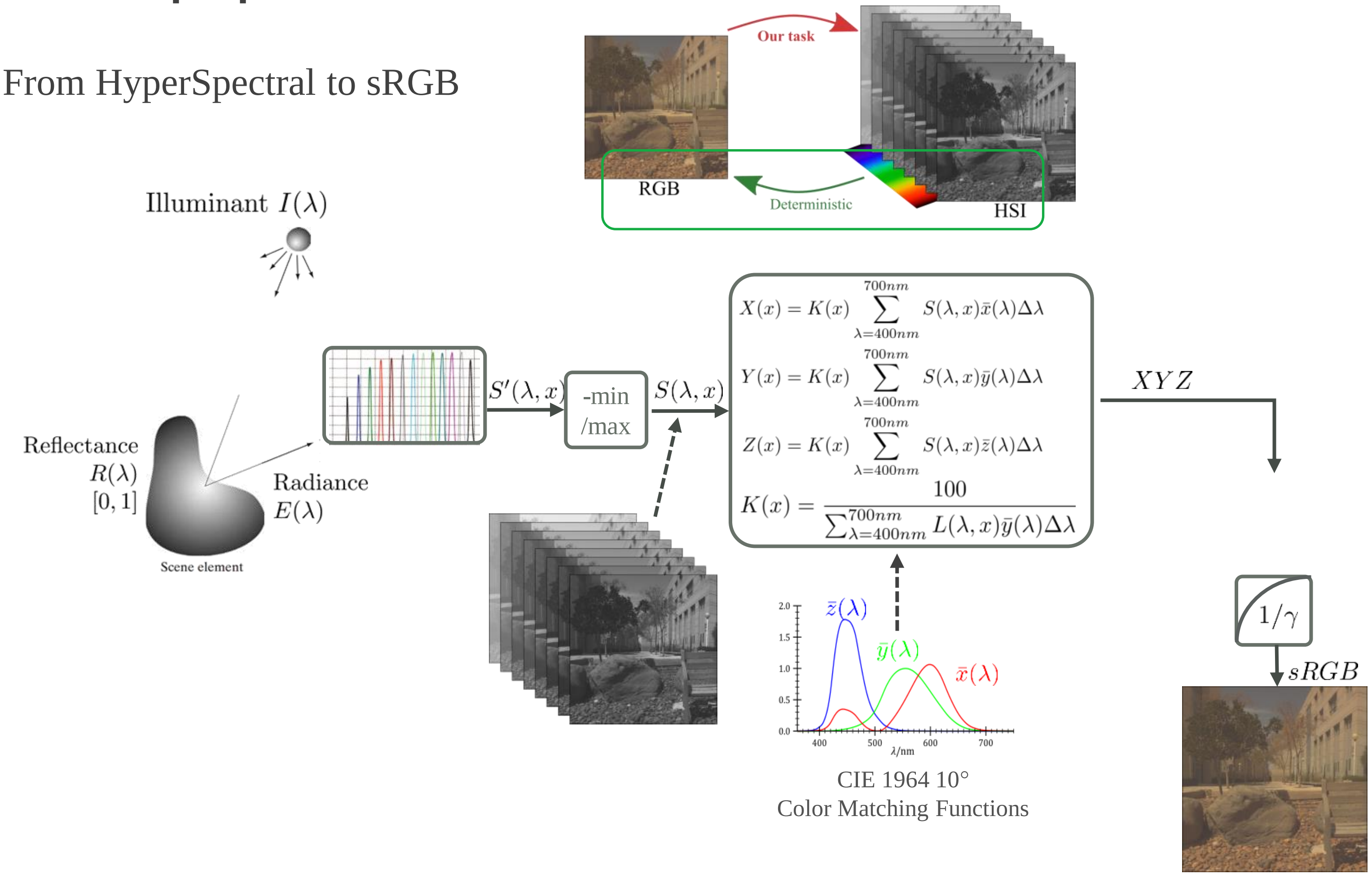
From HyperSpectral to sRGB





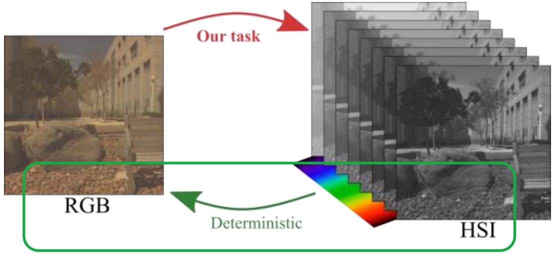
Dataset preparation

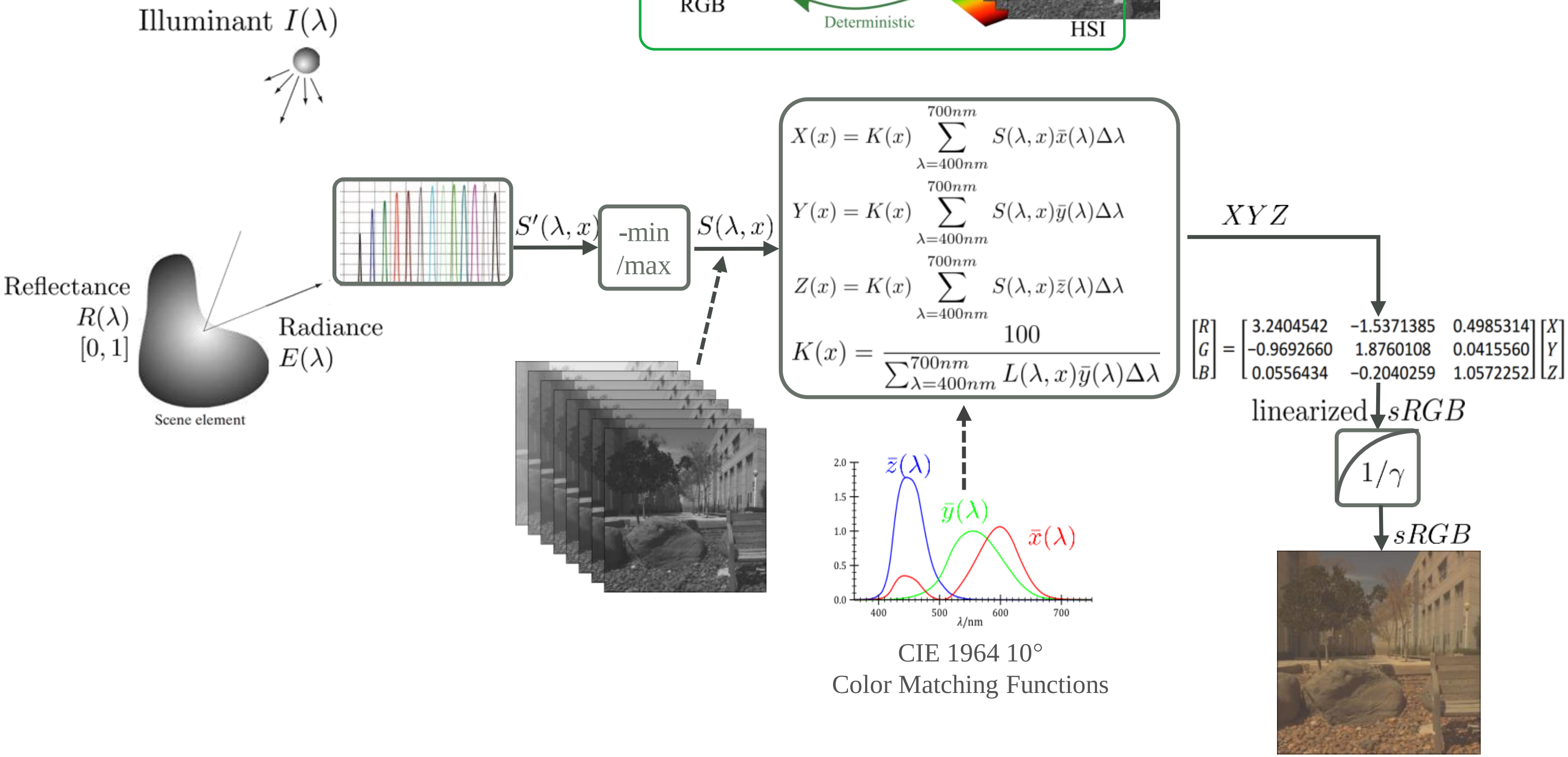
From HyperSpectral to sRGB



Dataset preparation

From HyperSpectral to sRGB





CIE 1964 10°
Color Matching Functions

Approach

3 Adversarial Networks for Spatial Context-Aware Spectral Image Reconstruction from RGB

RGB2HSI

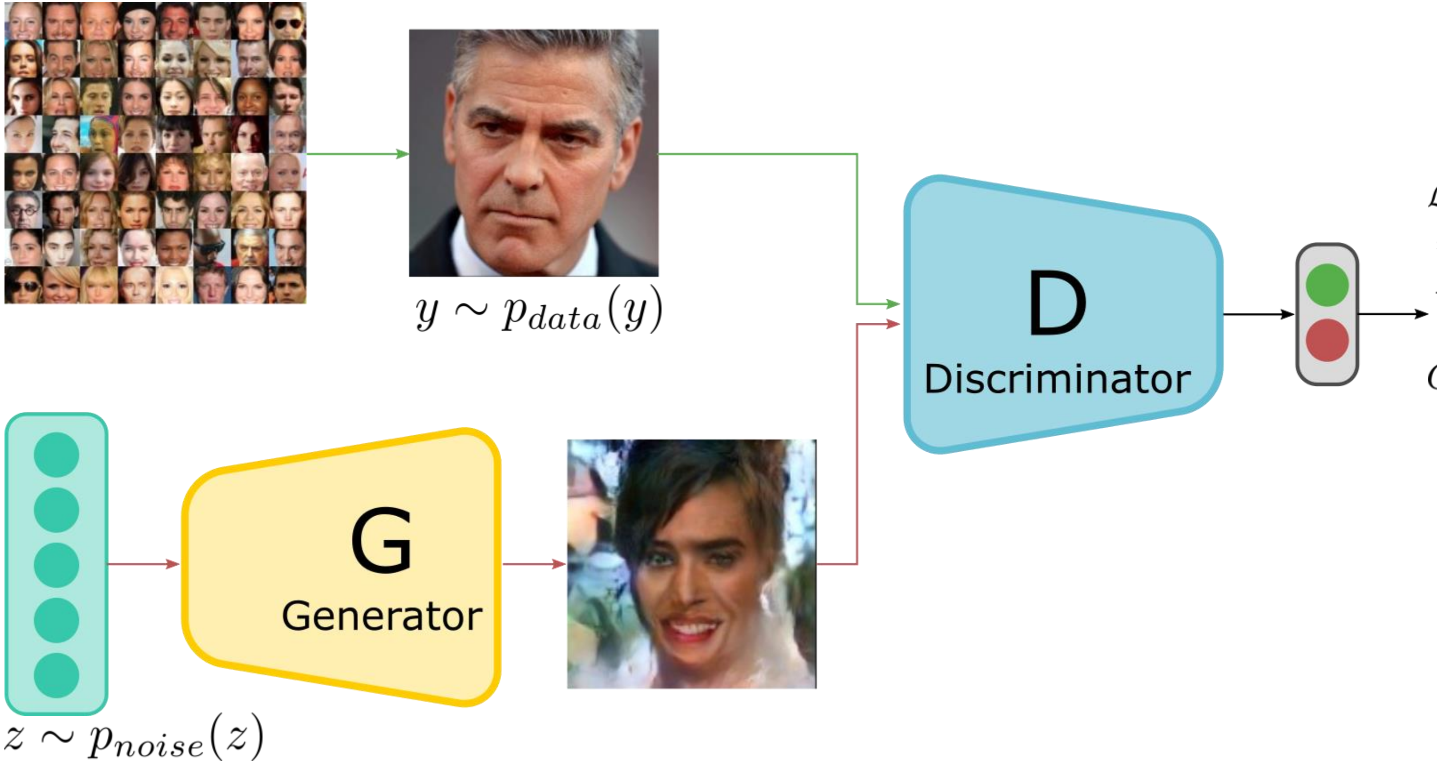
Approach

3 **Adversarial Networks** for Spatial Context-Aware Spectral Image Reconstruction from RGB

RGB2HSI

Generative adversarial Networks (GANs)

[Goodfellow2014]



$$\mathcal{L}_{GAN}(G, D) =$$
$$= \mathbb{E}_{y \sim p_{data}(y)}[\log D(y)] +$$
$$+ \mathbb{E}_{z \sim p_{noise}(z)}[\log(1 - D(G(z)))]$$

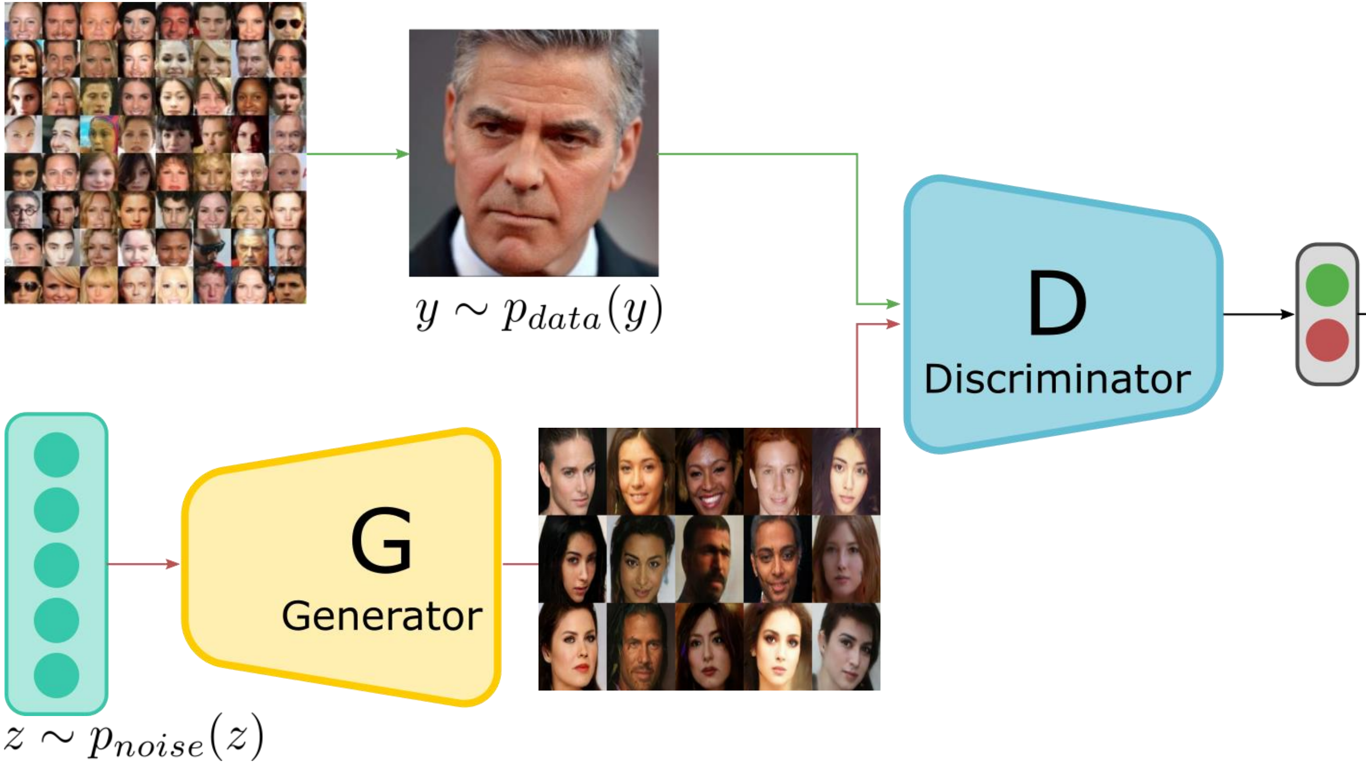
$$G^* = \arg\min_G \max_D \mathcal{L}_{GAN}$$

[Goodfellow2014] I. Goodfellow et al., "Generative Adversarial Nets," NIPS 2014

Sample images: D. Berthelot, "BEGAN: Boundary Equilibrium Generative Adversarial Networks," 2017. https://www.youtube.com/watch?v=J0o6LhaUSsc&vl=en

Generative adversarial Networks (GANs)

[Goodfellow2014]



$$\mathcal{L}_{GAN}(G, D) =$$
$$= \mathbb{E}_{y \sim p_{data}(y)}[\log D(y)] +$$
$$+ \mathbb{E}_{z \sim p_{noise}(z)}[\log(1 - D(G(z)))]$$

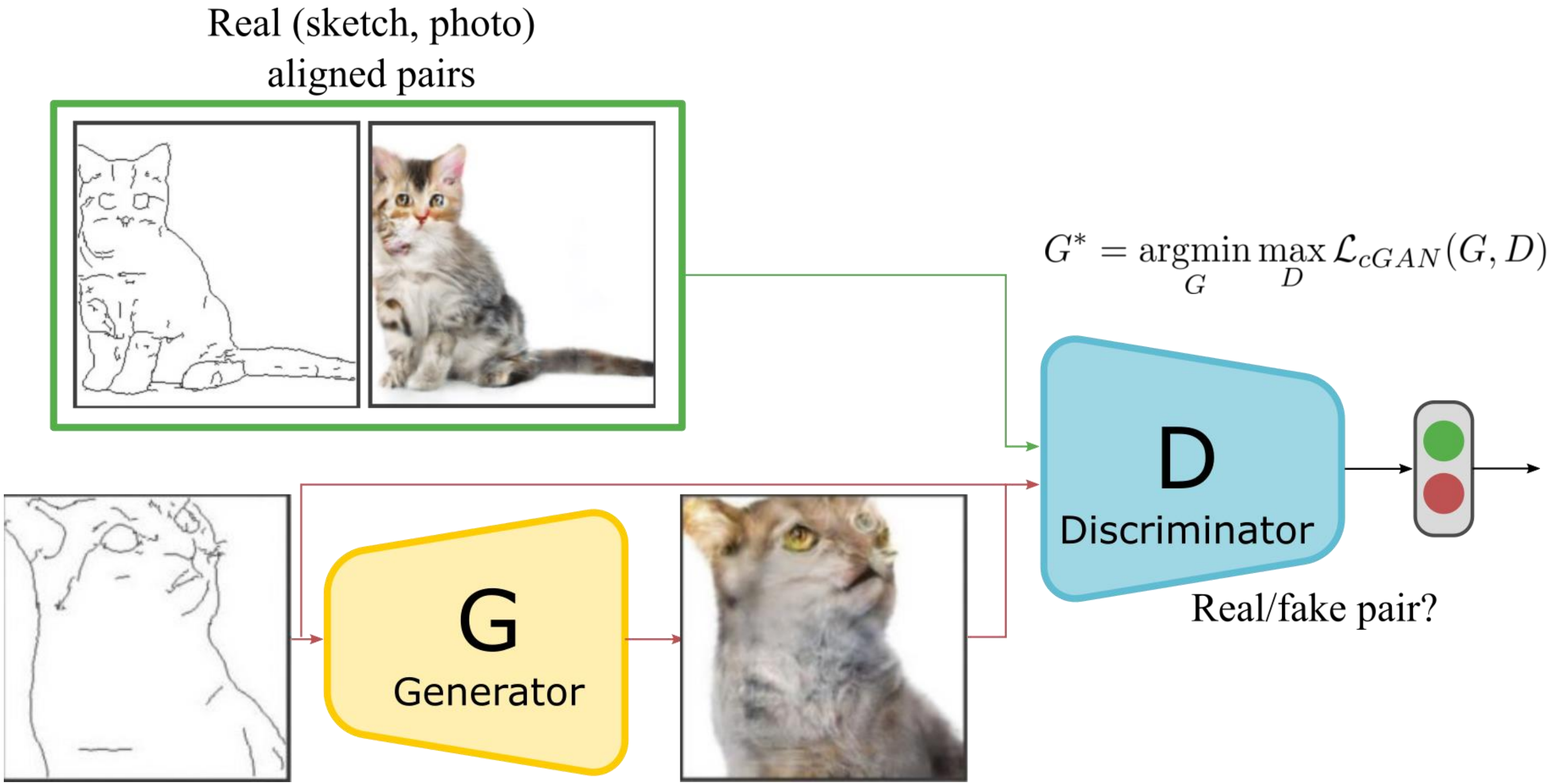
$$G^* = \arg\min_G \max_D \mathcal{L}_{GAN}$$

[Goodfellow2014] I. Goodfellow et al., "Generative Adversarial Nets," NIPS 2014

Sample images: D. Berthelot, "BEGAN: Boundary Equilibrium Generative Adversarial Networks," 2017. https://www.youtube.com/watch?v=J0o6LhaUSsc&vl=en

Conditional GANs: pix2pix

[Isola 2017]



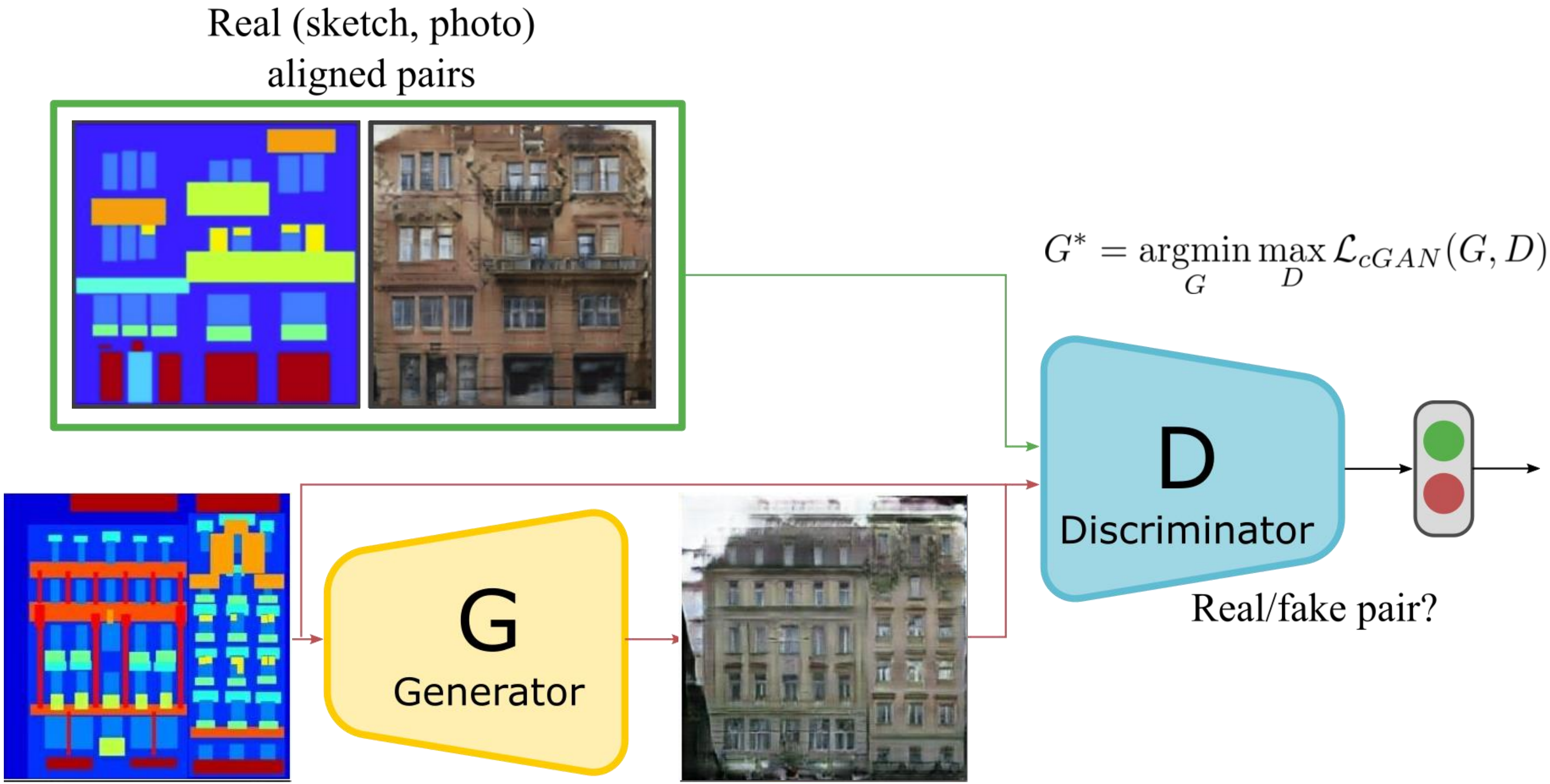
$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{\ell_1}(G)$$

[Isola2017] P. Isola et al., "Image-To-Image Translation With Conditional Adversarial Networks," CVPR 2017.

[Hesse2017] C. Hesse, "Image-to-Image Demo - Affine Layer," 2017, https://affinelayer.com/pixsrv/.

Conditional GANs: pix2pix

[Isola 2017]



$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{\ell_1}(G)$$

[Isola2017] P. Isola et al., "Image-To-Image Translation With Conditional Adversarial Networks," CVPR 2017.

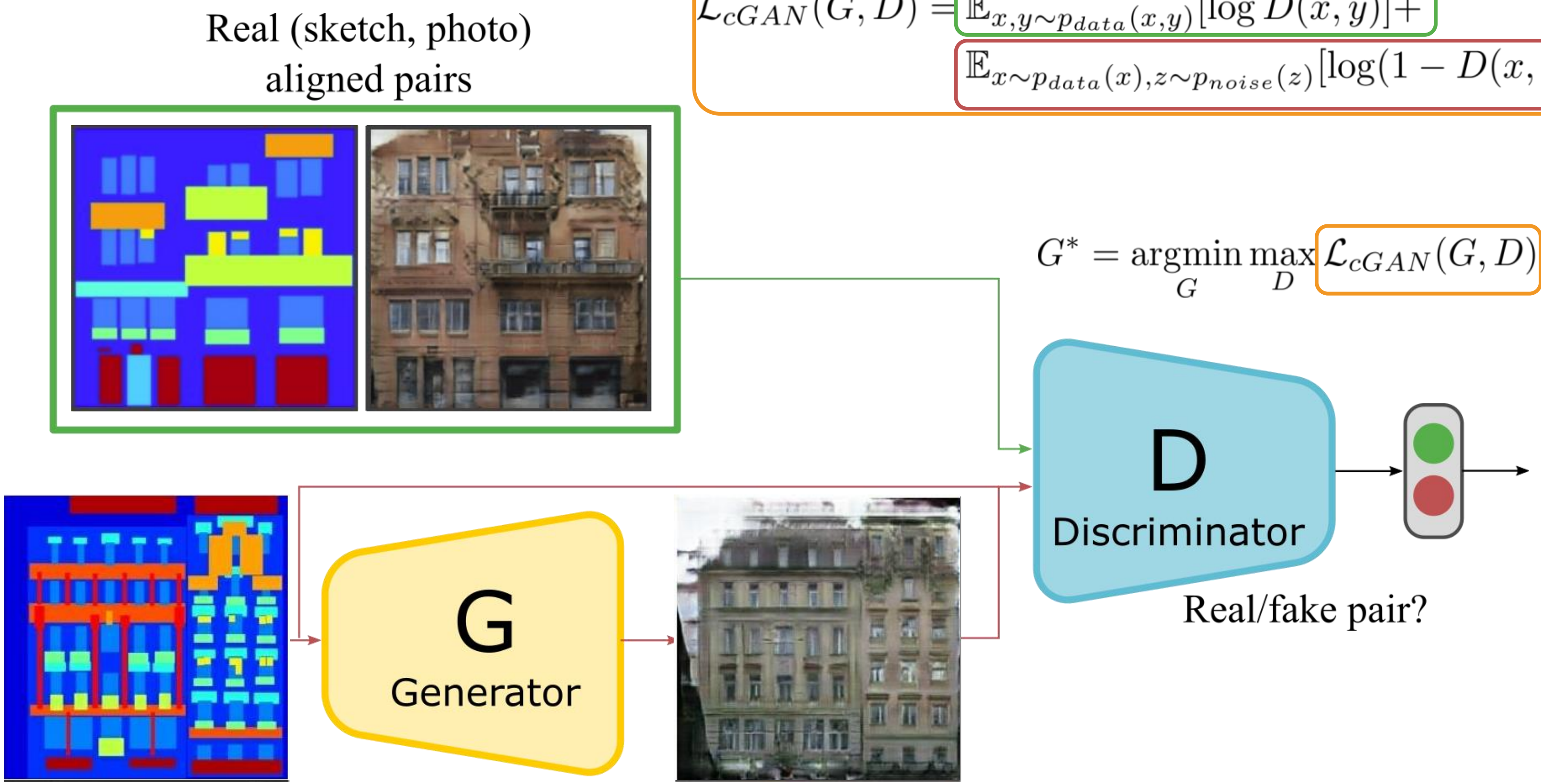
[Hesse2017] C. Hesse, "Image-to-Image Demo - Affine Layer," 2017, https://affinelayer.com/pixsrv/.

Conditional GANs: pix2pix

[Isola 2017]

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x,y)] + \mathbb{E}_{x\sim p_{data}(x),z\sim p_{noise}(z)}[\log(1-D(x,G(x,z)))]$$

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda \mathcal{L}_{\ell_1}(G)$$

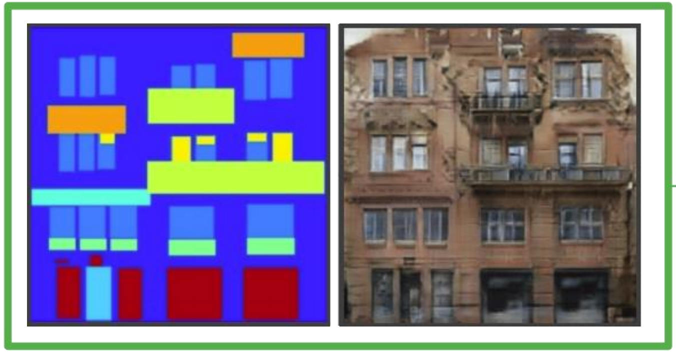


[Isola2017] P. Isola et al., "Image-To-Image Translation With Conditional Adversarial Networks," CVPR 2017.
[Hesse2017] C. Hesse, "Image-to-Image Demo - Affine Layer," 2017, https://affinelayer.com/pixsrv/.

Conditional GANs: pix2pix

[Isola 2017]

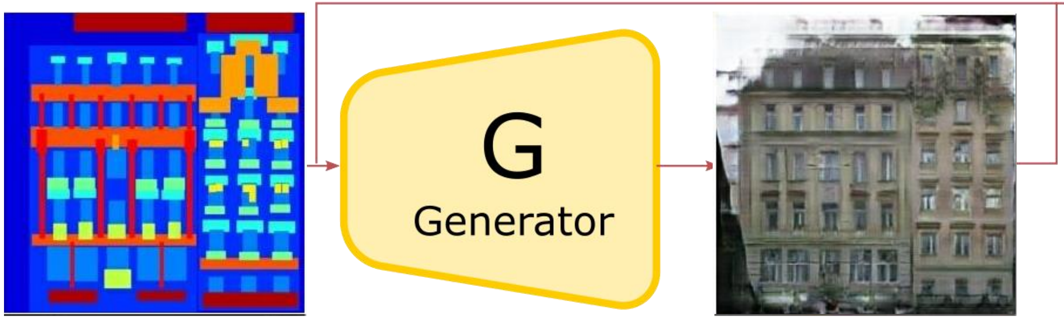
Real (sketch, photo)
aligned pairs

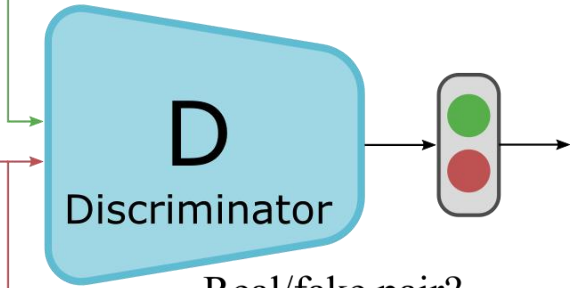


$$\mathcal{L}_{cGAN}(G,D) = \boxed{\mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x,y)]+} \\ \boxed{\mathbb{E}_{x\sim p_{data}(x),z\sim p_{noise}(z)}[\log(1-D(x,G(x,z)))]}$$

$$\boxed{\mathcal{L}_{\ell_1}(G) = \mathbb{E}_{x,y\sim p_{data}(x,y),z\sim p_{noise}(z)}[\|y-G(x,z)\|_1]}$$

$$G^* = \arg\min_G \max_D \boxed{\mathcal{L}_{cGAN}(G,D)} + \lambda \boxed{\mathcal{L}_{\ell_1}(G)}$$





Real/fake pair?

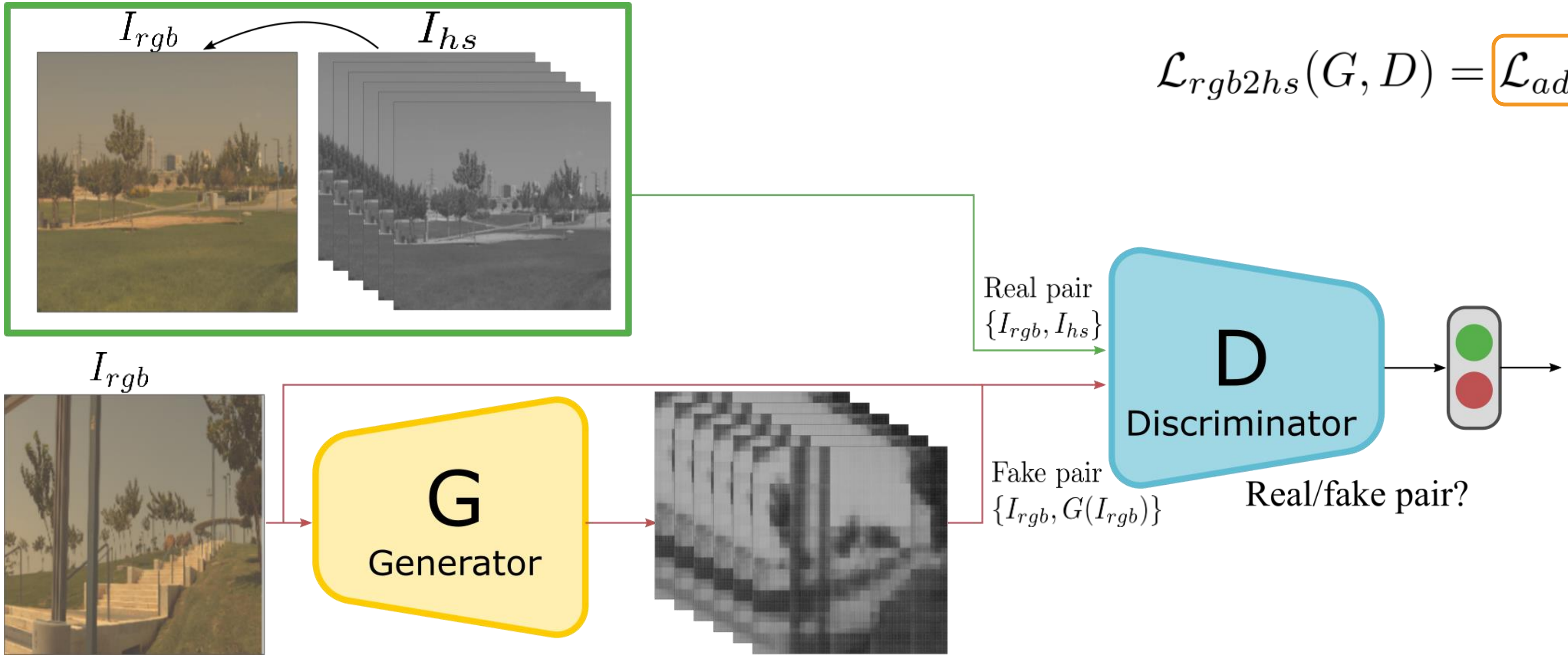
[Isola2017] P. Isola et al., "Image-To-Image Translation With Conditional Adversarial Networks," CVPR 2017.
[Hesse2017] C. Hesse, "Image-to-Image Demo - Affine Layer," 2017, https://affinelayer.com/pixsrv/.

Adversarial spectral reconstruction networks

$$\mathcal{L}_{adv}(G,D) = \boxed{\mathbb{E}_{I_{rgb},I_{hs}\sim p_{data}(I_{rgb},I_{hs})}[\log D(I_{rgb},I_{hs})]} + \boxed{\mathbb{E}_{I_{rgb}\sim p_{data}(I_{rgb})}[\log(1-D(I_{rgb},G(I_{rgb})))]}$$

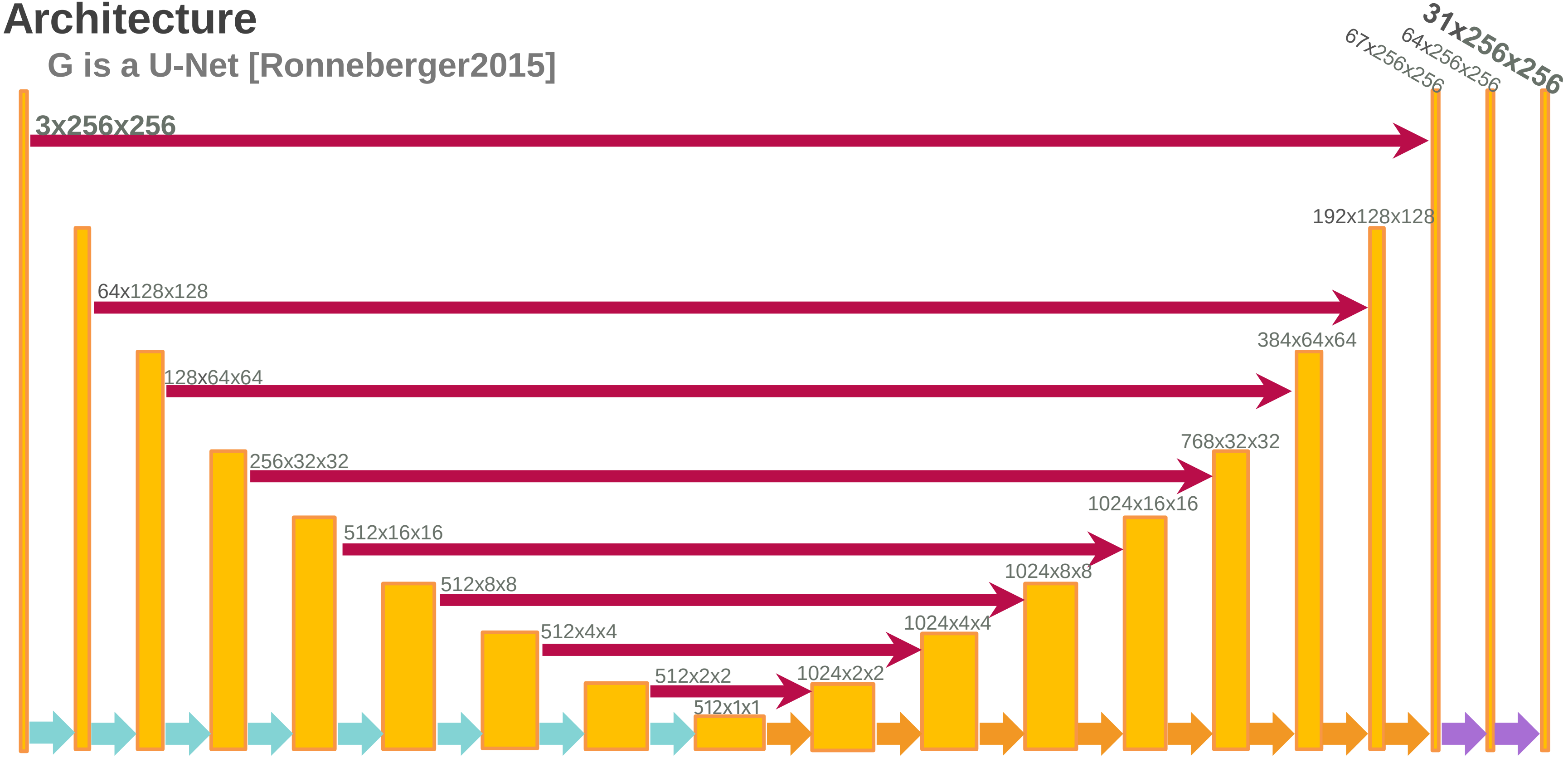
$$\mathcal{L}_{\ell_1} = \mathbb{E}_{I_{rgb},I_{hs}\sim p_{data}(I_{rgb},I_{hs})}[\|I_{hs}-G(I_{rgb})\|_1]$$

$$\mathcal{L}_{rgb2hs}(G,D) = \boxed{\mathcal{L}_{adv}} + \lambda\boxed{\mathcal{L}_{\ell_1}}$$



Architecture

G is a U-Net [Ronneberger2015]



Conv2D(k=3,s=2) + BatchNorm + LeakyReLU

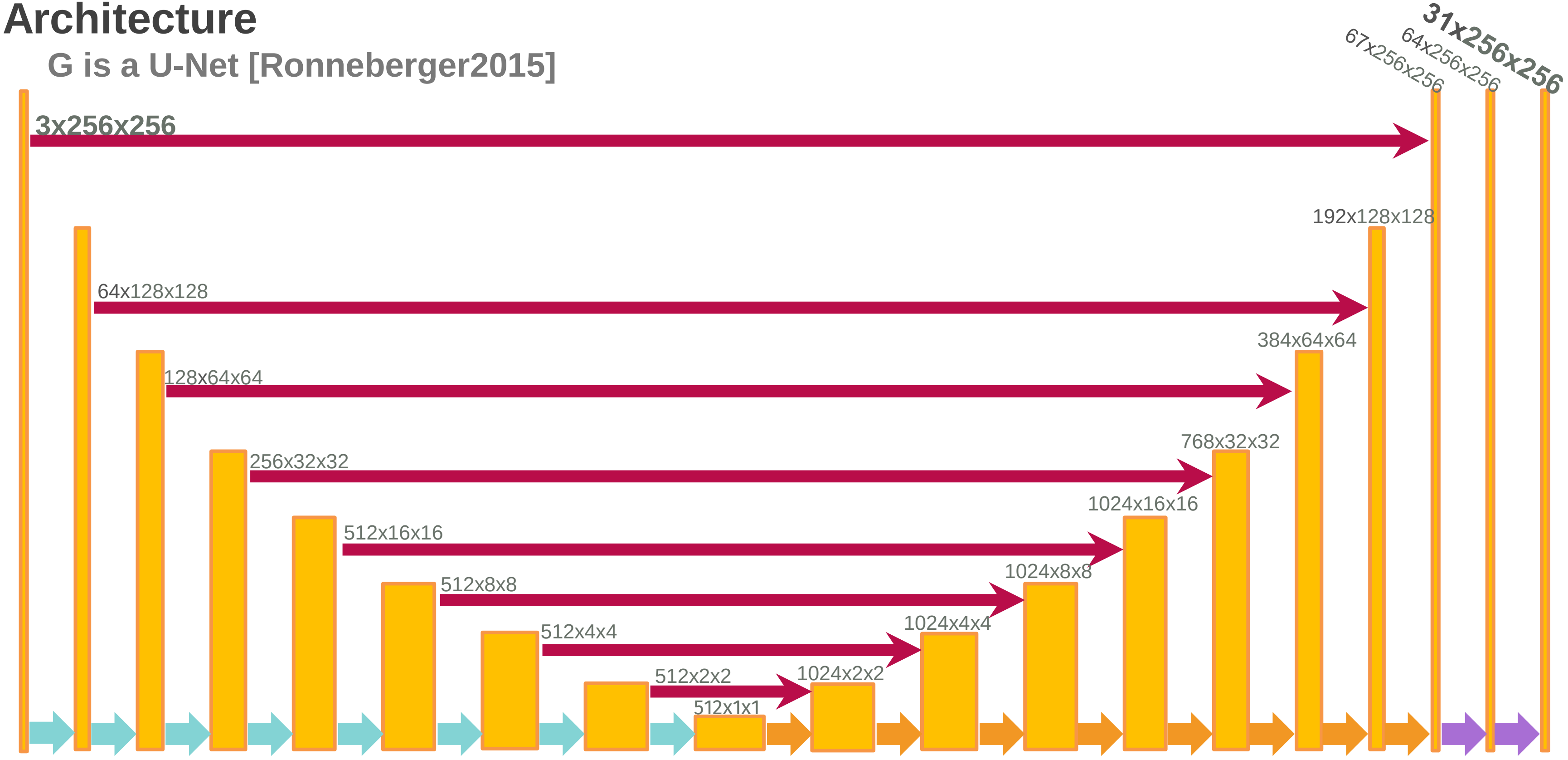
Conv2DTranspose(k=2,s=1) + Dropout(r=0.1) + Merge( ) + LeakyReLU

Conv2D(k=1,s=1) + LeakyReLU

[Ronneberger2015] O. Ronneberger et al., "*U-Net: Convolutional Networks for Biomedical Image Segmentation*," MICCAI 2015

Architecture

G is a U-Net [Ronneberger2015]



Conv2D(k=3,s=2) + ~~BatchNorm~~ + LeakyReLU

Conv2DTranspose(k=2,s=1) + Dropout(r=0.1) + Merge( ) + LeakyReLU

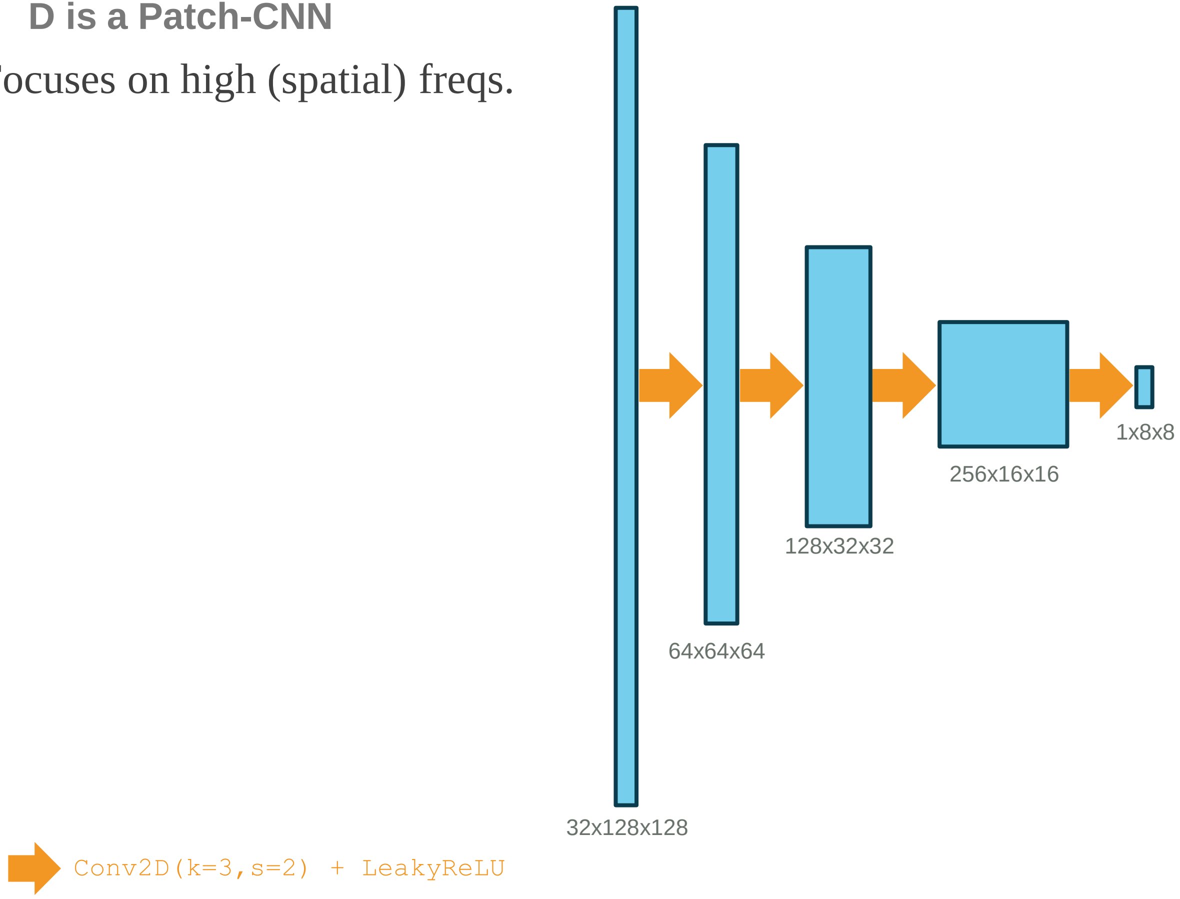
Conv2D(k=1,s=1) + LeakyReLU

[Ronneberger2015] O. Ronneberger et al., "*U-Net: Convolutional Networks for Biomedical Image Segmentation*," MICCAI 2015

Architecture

D is a Patch-CNN

Focuses on high (spatial) freqs.



Experiments

Evaluation on ICVL

Per pixel, across the spectrum

Normalized by radiance
Accounts for low luminance samples

Goodness of fit coefficient (high is better)

Perceptual color difference

Method	RMSE	RMSERel	GFC	ΔE_{00}
[Arad2016] reported	2.633	0.0756	-	-
[Arad2016] optimized	2.184 ± 0.064	0.0872 ± 0.004	—	—
[Galliani2017] reported	1.980	0.0587	—	—
Ours	$\mathbf{1.457 \pm 0.040}$	$\mathbf{0.0401 \pm 0.0024}$	$\mathbf{0.99921 \pm 0.00012}$	$\mathbf{2.044 \pm 0.341}$
fold 0	1.452 ± 0.101	0.0383 ± 0.0024	0.99906 ± 0.00001	1.861 ± 0.324
fold 1	1.463 ± 0.022	0.0420 ± 0.0024	0.99936 ± 0.00023	2.228 ± 0.358

Experiments

Evaluation on ICVL

Per pixel, across the spectrum

Normalized by radiance Accounts for low luminance samples

Goodness of fit coefficient (high is better)

Perceptual color difference

Method	RMSE	RMSERel	GFC	ΔE_{00}
[Arad2016] reported	2.633	0.0756	-	-
[Arad2016] optimized	2.184 ± 0.064 -33.2%	0.0872 ± 0.004 -54,0%	—	—
[Galliani2017] reported	1.980	0.0587	—	—
Ours	$\mathbf{1.457 \pm 0.040}$	$\mathbf{0.0401 \pm 0.0024}$	$\mathbf{0.99921 \pm 0.00012}$	$\mathbf{2.044 \pm 0.341}$
fold 0	1.452 ± 0.101	0.0383 ± 0.0024	0.99906 ± 0.00001	1.861 ± 0.324
fold 1	1.463 ± 0.022	0.0420 ± 0.0024	0.99936 ± 0.00023	2.228 ± 0.358

Pixelwise reconstruction

Experiments

Evaluation on ICVL

Per pixel, across the spectrum

Normalized by radiance
Accounts for low luminance samples

Goodness of fit coefficient (high is better)

Perceptual color difference

Method	RMSE	RMSERel	GFC	ΔE_{00}
[Arad2016] reported	2.633	0.0756	-	-
[Arad2016] optimized	2.184 ± 0.064 -33.2%	0.0872 ± 0.004 -54,0%	–	–
[Galliani2017] reported	1.980	0.0587	–	–
Ours	$\mathbf{1.457 \pm 0.040}$	$\mathbf{0.0401 \pm 0.0024}$	$\mathbf{0.99921 \pm 0.00012}$	$\mathbf{2.044 \pm 0.341}$
fold 0	1.452 ± 0.101	0.0383 ± 0.0024	0.99906 ± 0.00001	1.861 ± 0.324
fold 1	1.463 ± 0.022	0.0420 ± 0.0024	0.99936 ± 0.00023	2.228 ± 0.358

Pixelwise reconstruction

CNN

Experiments

Evaluation on ICVL

Per pixel, across the spectrum

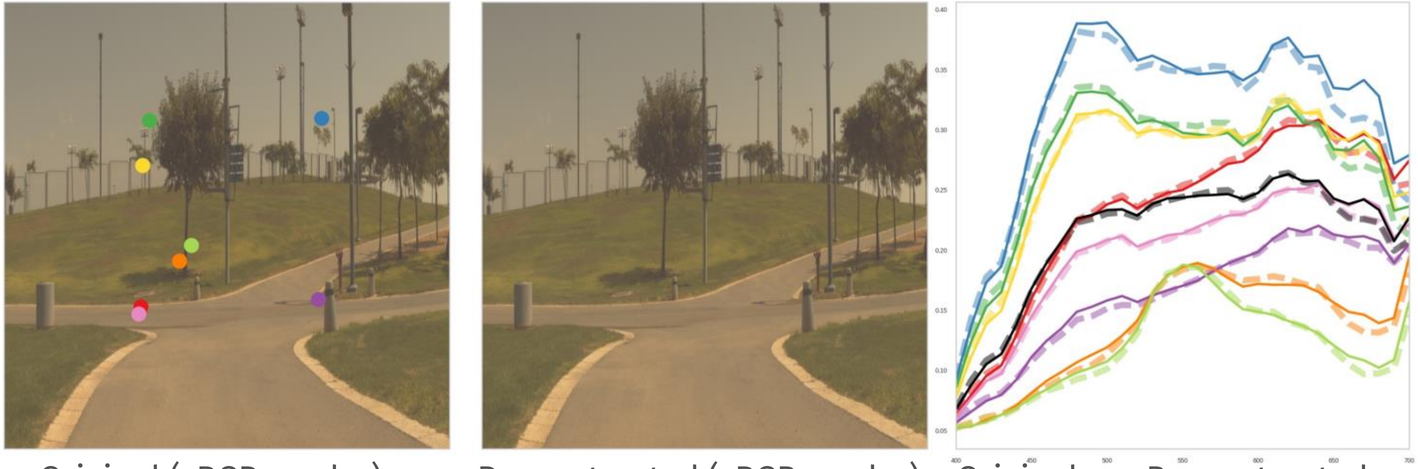
Normalized by radiance Accounts for low luminance samples

Goodness of fit coefficient (high is better)

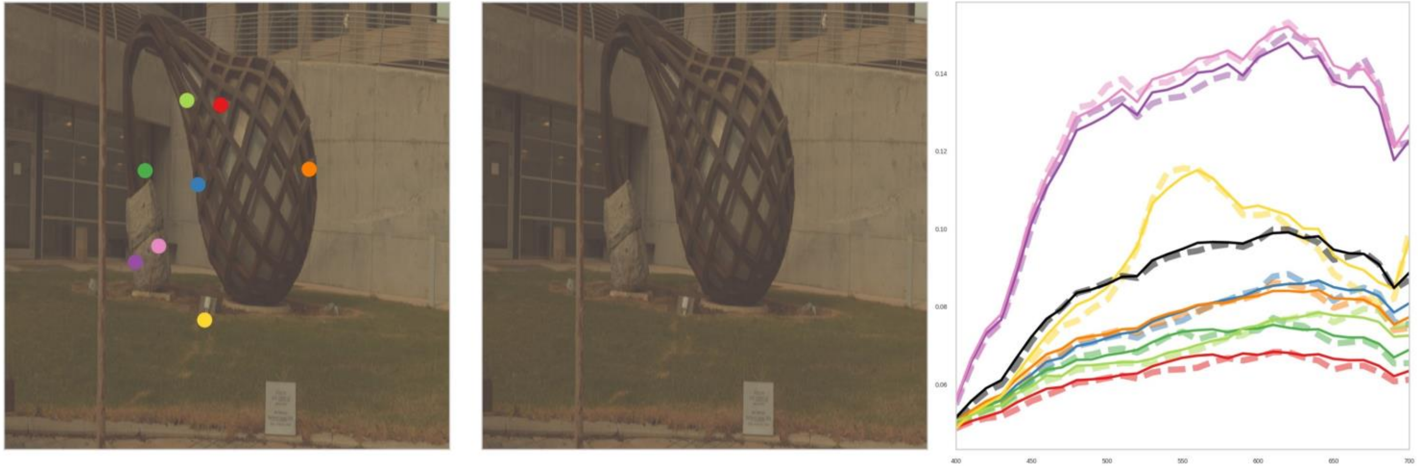
Perceptual color difference

Method	RMSE	RMSERel	GFC	ΔE_{00}	
[Arad2016] reported	2.633	0.0756	-	-	Pixelwise reconstruction
[Arad2016] optimized	2.184 ± 0.064 -33.2%	0.0872 ± 0.004 -54,0%	—	—	
[Galliani2017] reported	1.980	0.0587	—	—	CNN
Ours	$\mathbf{1.457 \pm 0.040}$	$\mathbf{0.0401 \pm 0.0024}$	$\mathbf{0.99921 \pm 0.00012}$	$\mathbf{2.044 \pm 0.341}$	
fold 0	1.452 ± 0.101	0.0383 ± 0.0024	0.99906 ± 0.00001	1.861 ± 0.324	
fold 1	1.463 ± 0.022	0.0420 ± 0.0024	0.99936 ± 0.00023	2.228 ± 0.358	

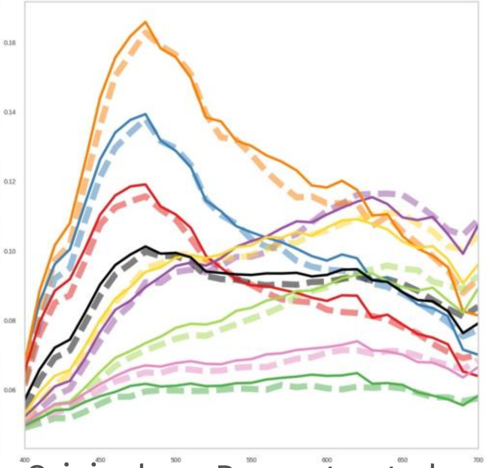
Results



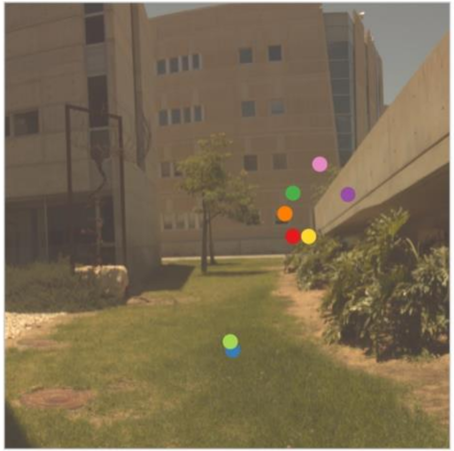
Original (sRGB render) Reconstructed (sRGB render) Original --- Reconstructed —

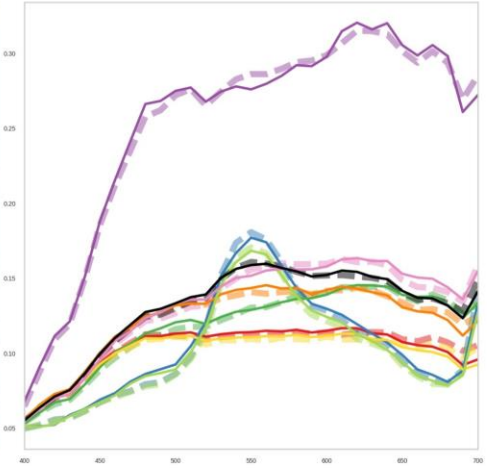


Results

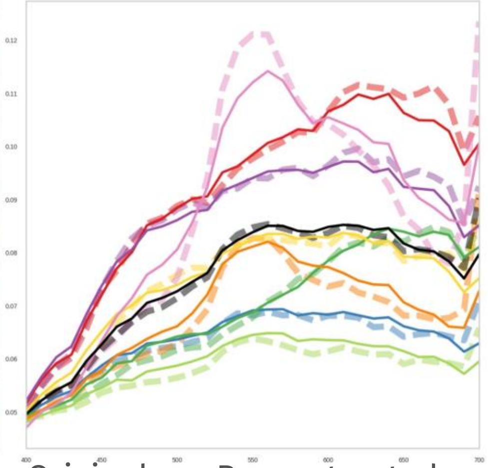


Original (sRGB render) Reconstructed (sRGB render) Original --- Reconstructed —

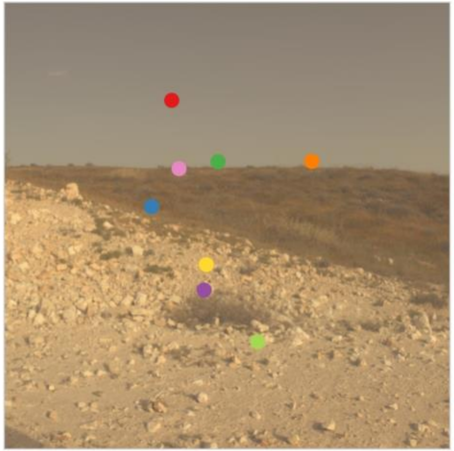


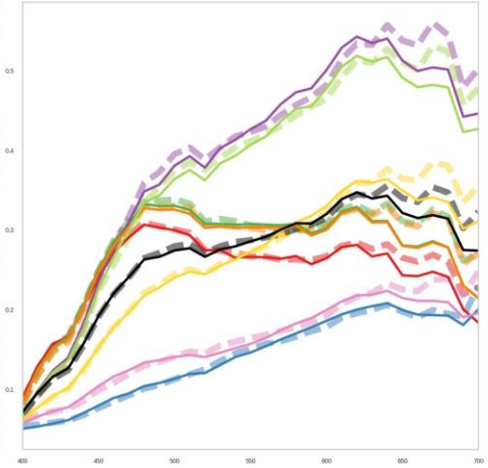


Results



Original (sRGB render) Reconstructed (sRGB render) Original --- Reconstructed —





Results

3 Adversarial Networks for Spatial Context-Aware Spectral Image Reconstruction from RGB

RGB2HSI

Results

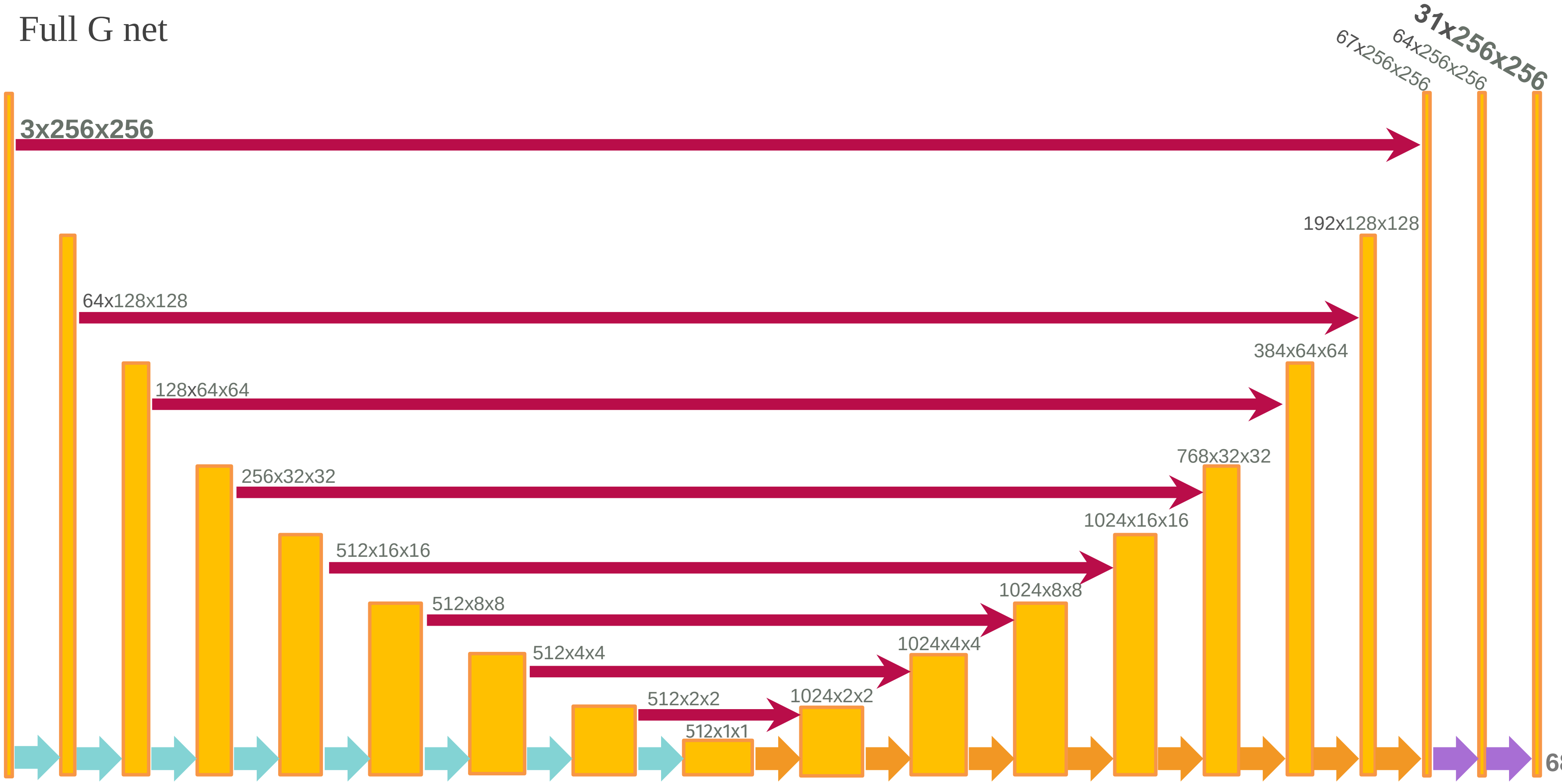
3 Adversarial Networks for Spatial Context-Aware Spectral Image Reconstruction from RGB

RGB2HSI

Results

Does spatial information really help?

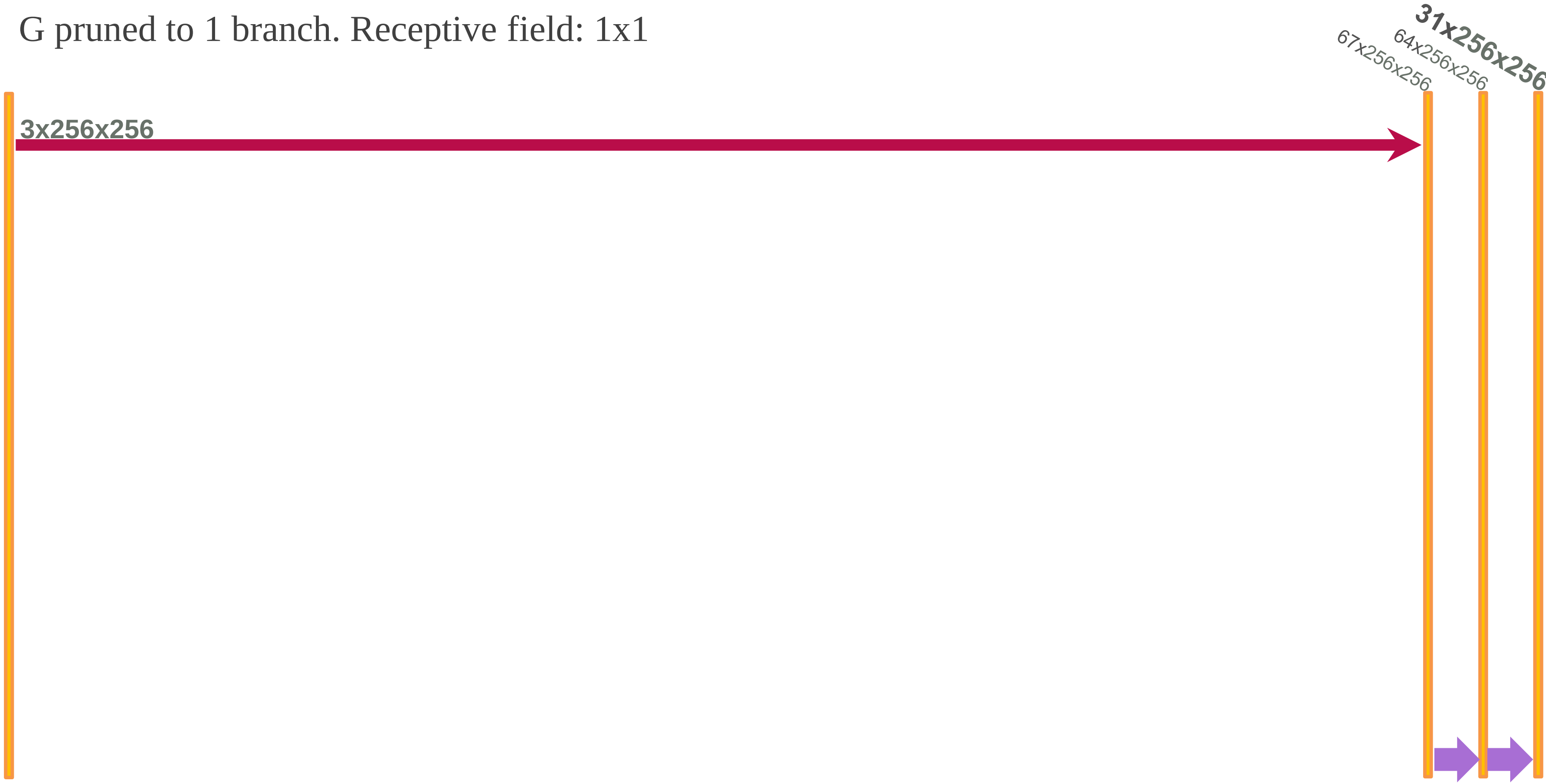
Full G net



Results

Does spatial information really help?

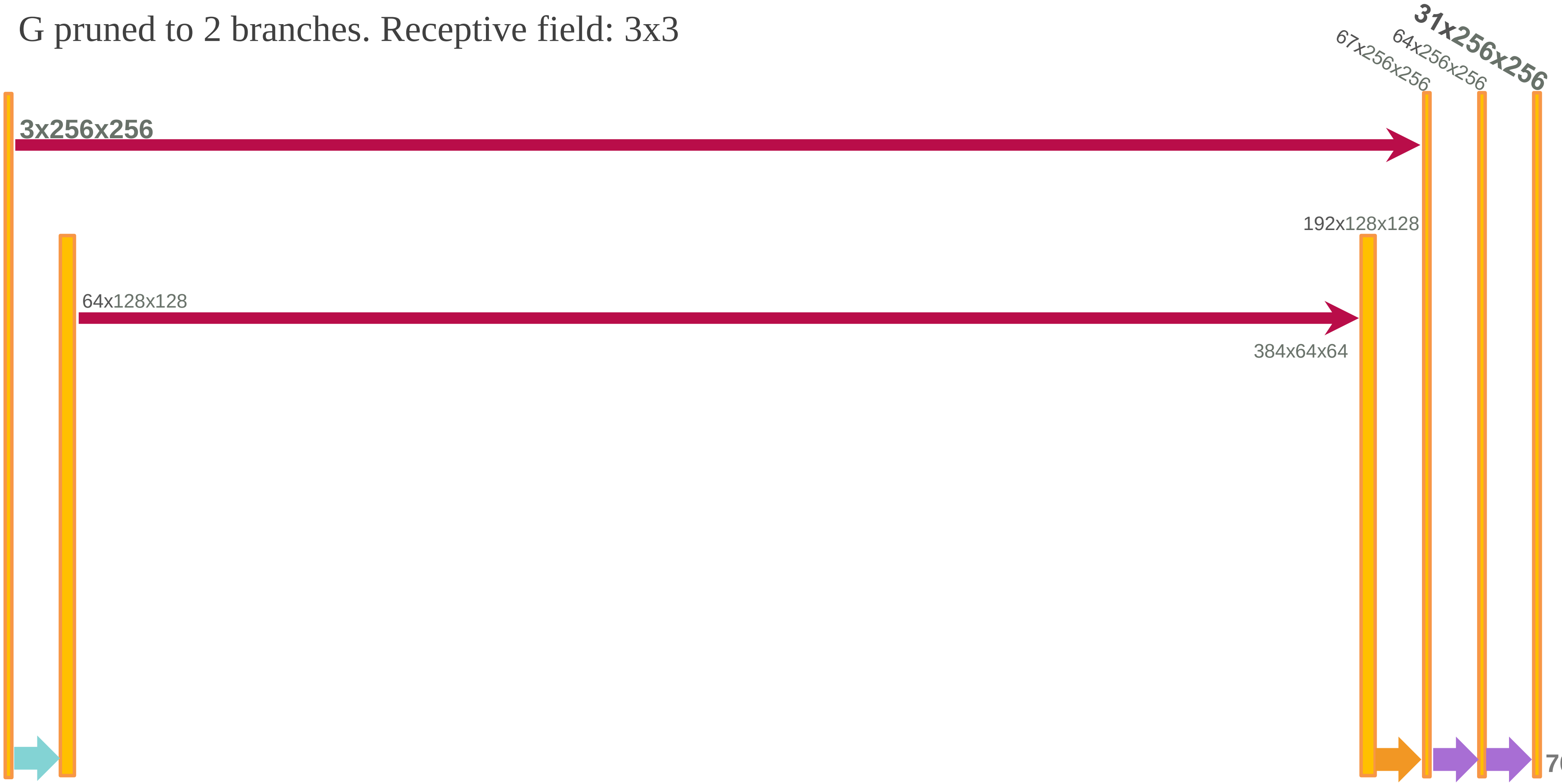
G pruned to 1 branch. Receptive field: 1x1



Results

Does spatial information really help?

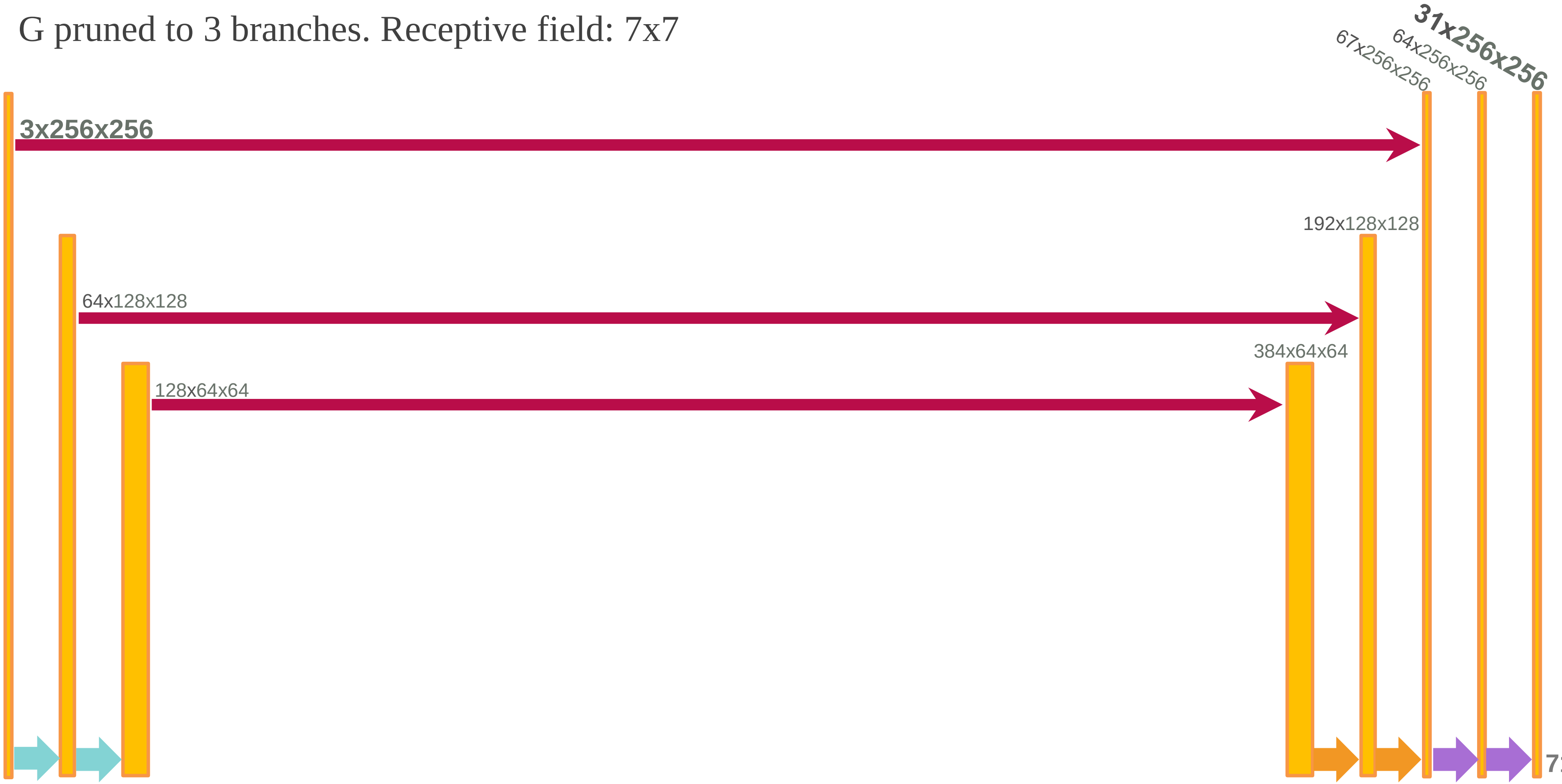
G pruned to 2 branches. Receptive field: 3x3



Results

Does spatial information really help?

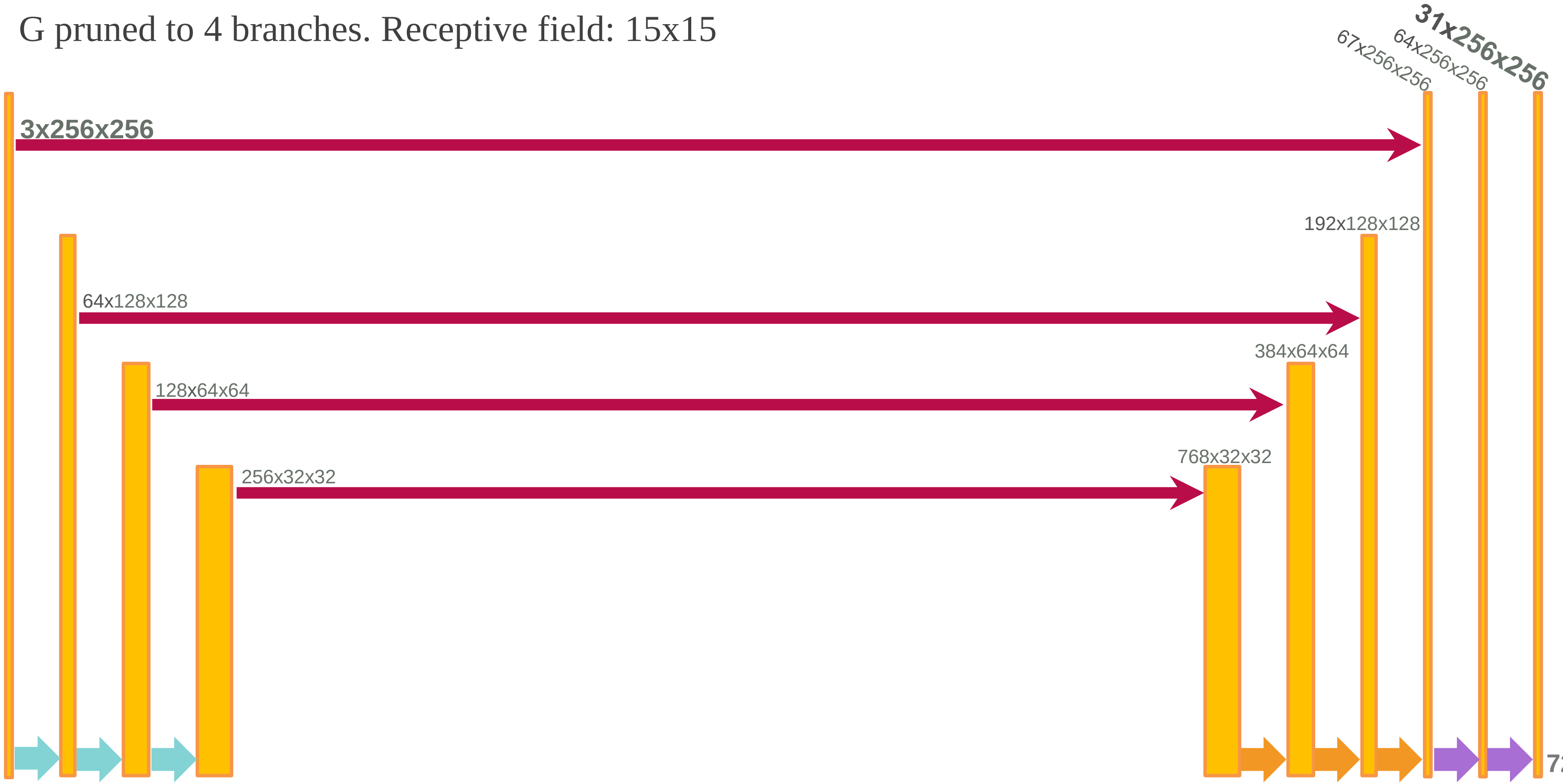
G pruned to 3 branches. Receptive field: 7x7



Results

Does spatial information really help?

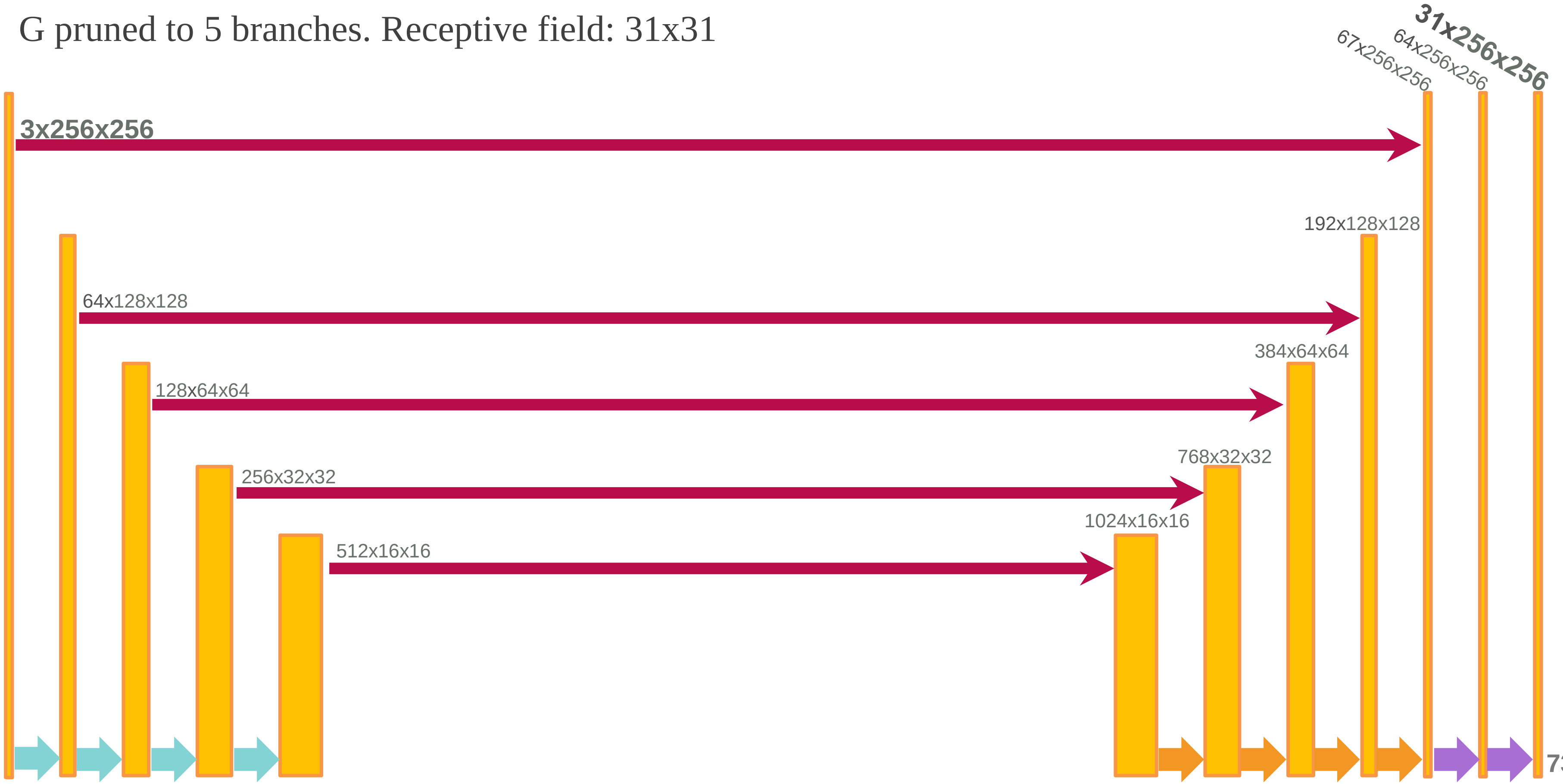
G pruned to 4 branches. Receptive field: 15x15



Results

Does spatial information really help?

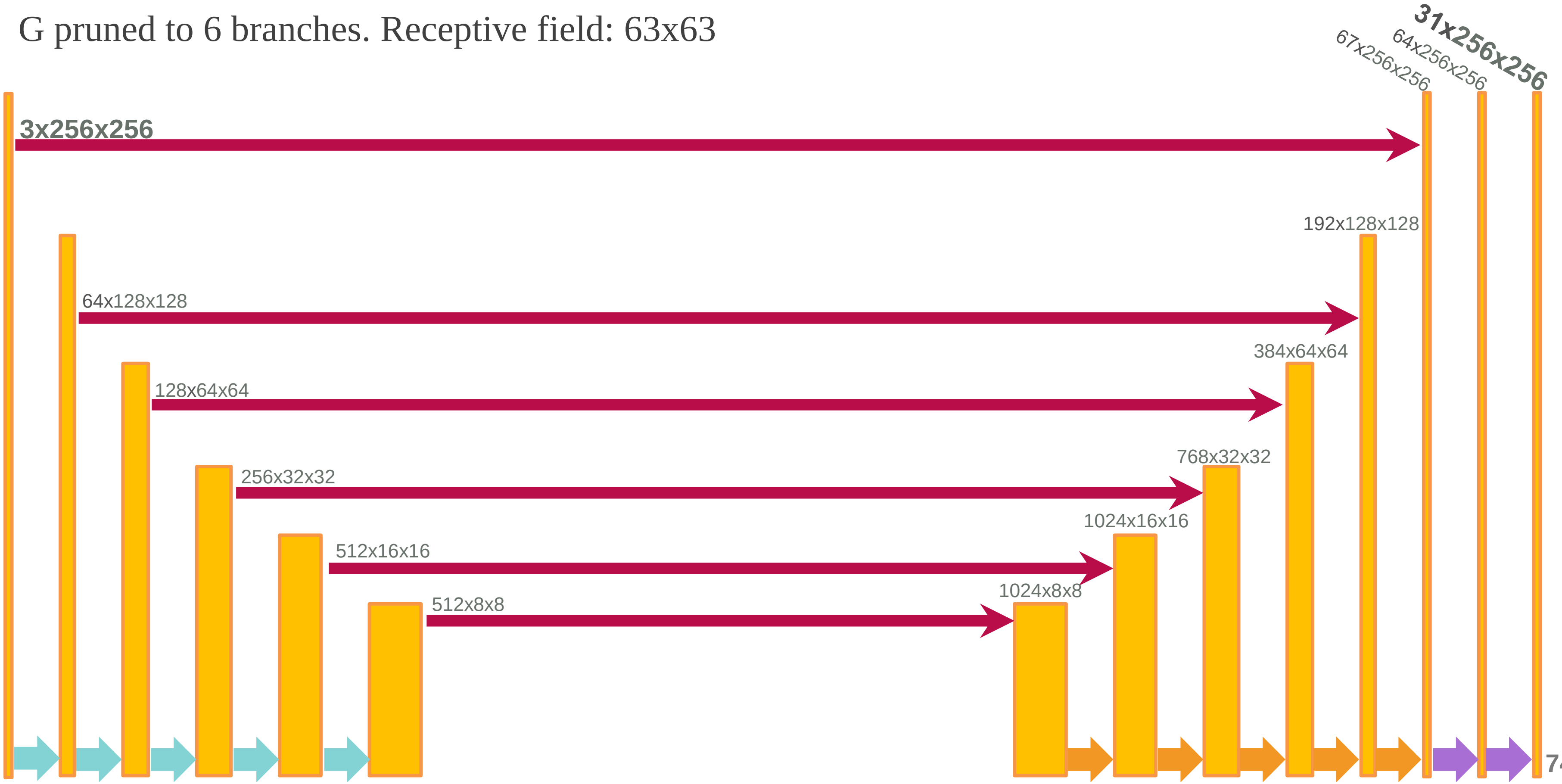
G pruned to 5 branches. Receptive field: 31x31



Results

Does spatial information really help?

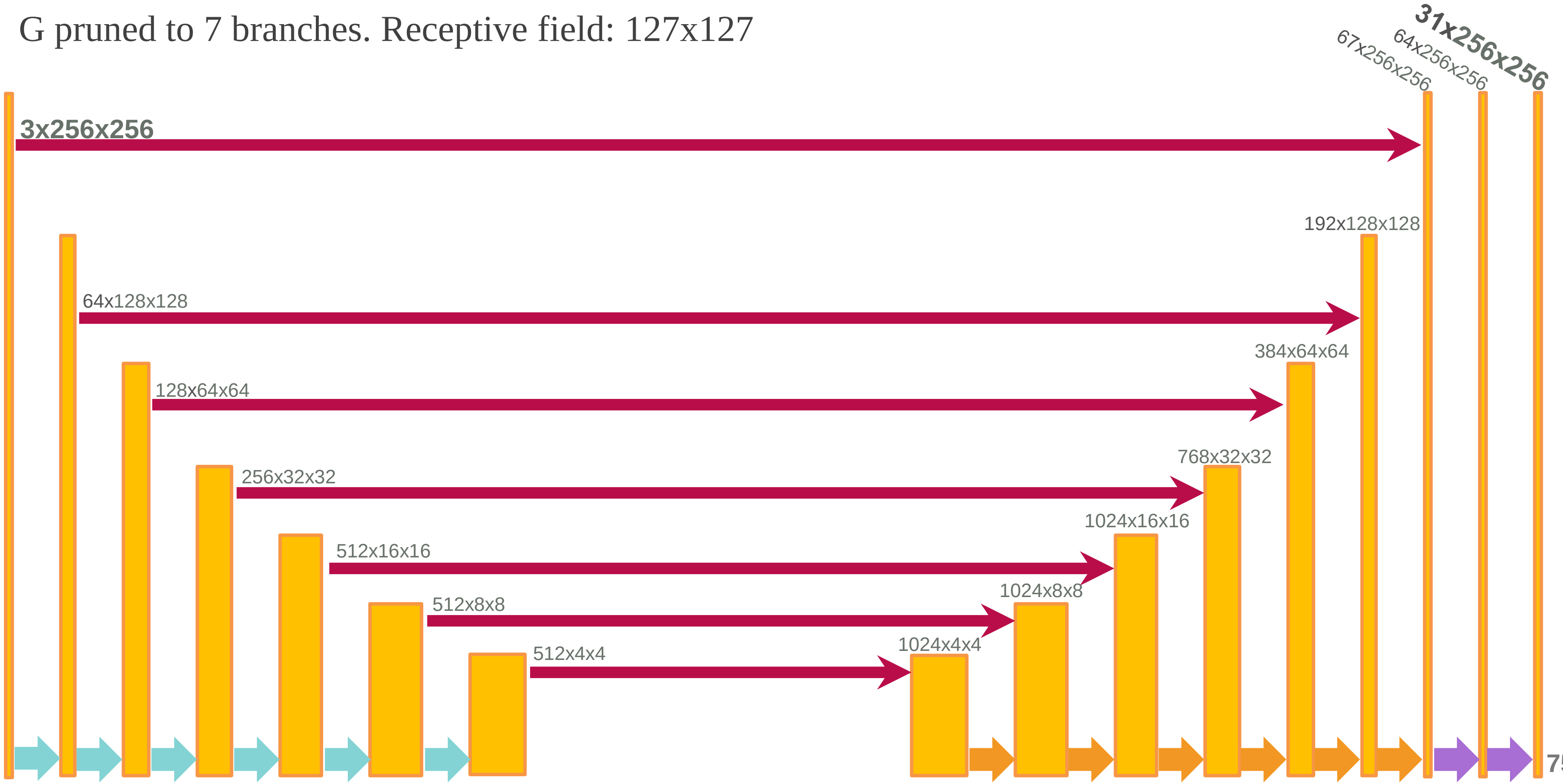
G pruned to 6 branches. Receptive field: 63x63



Results

Does spatial information really help?

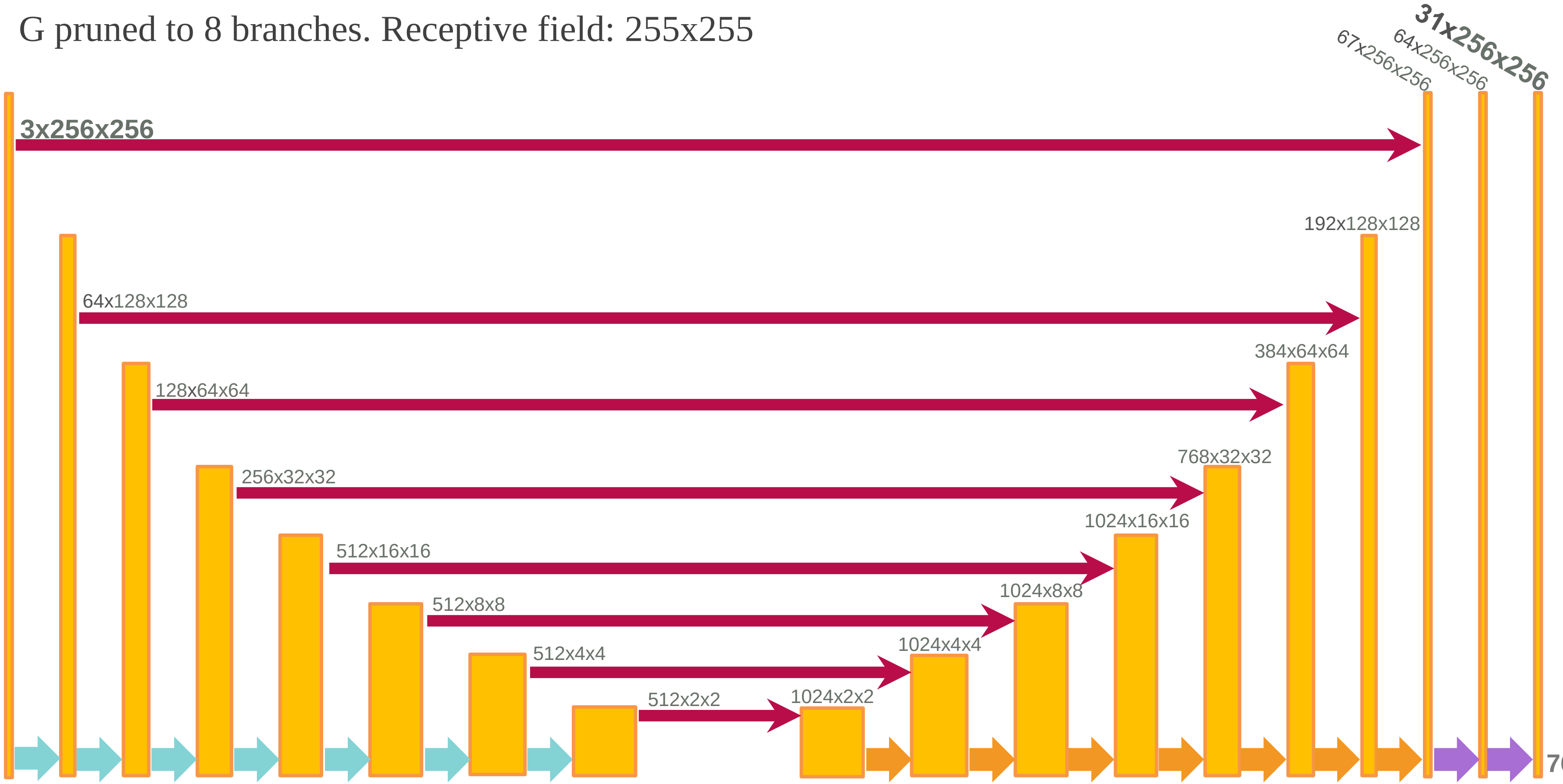
G pruned to 7 branches. Receptive field: 127x127



Results

Does spatial information really help?

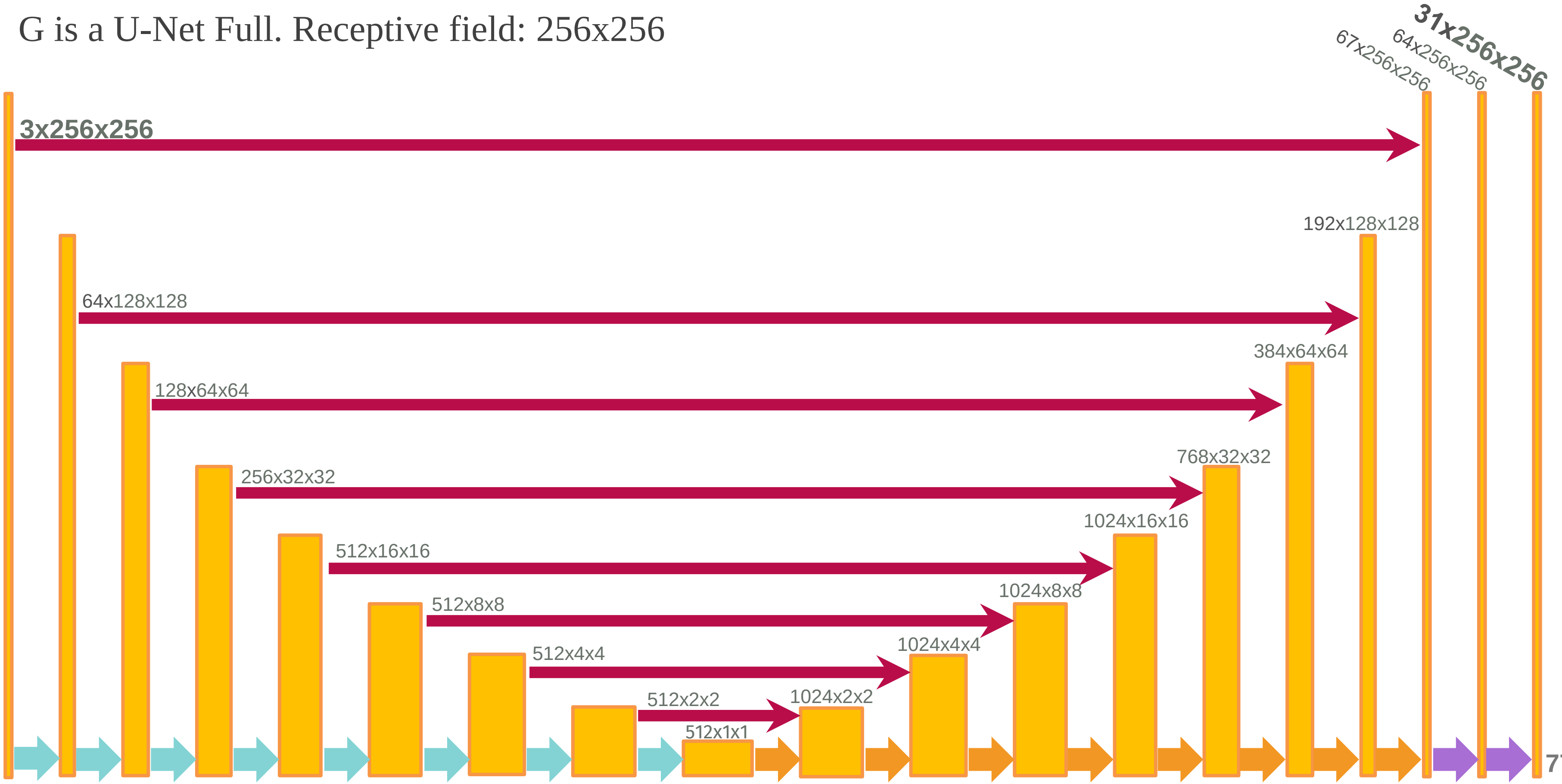
G pruned to 8 branches. Receptive field: 255x255



Results

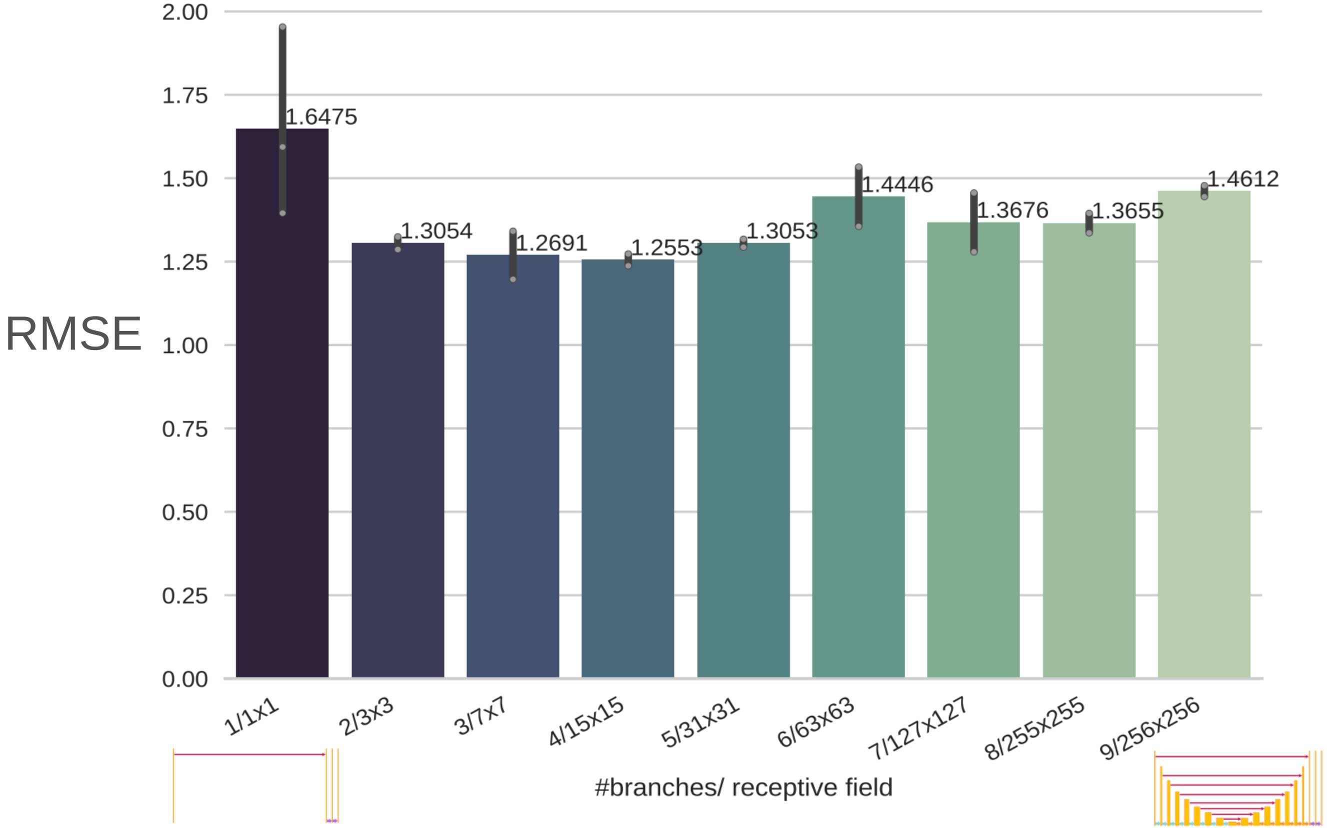
Does spatial information really help?

G is a U-Net Full. Receptive field: 256x256



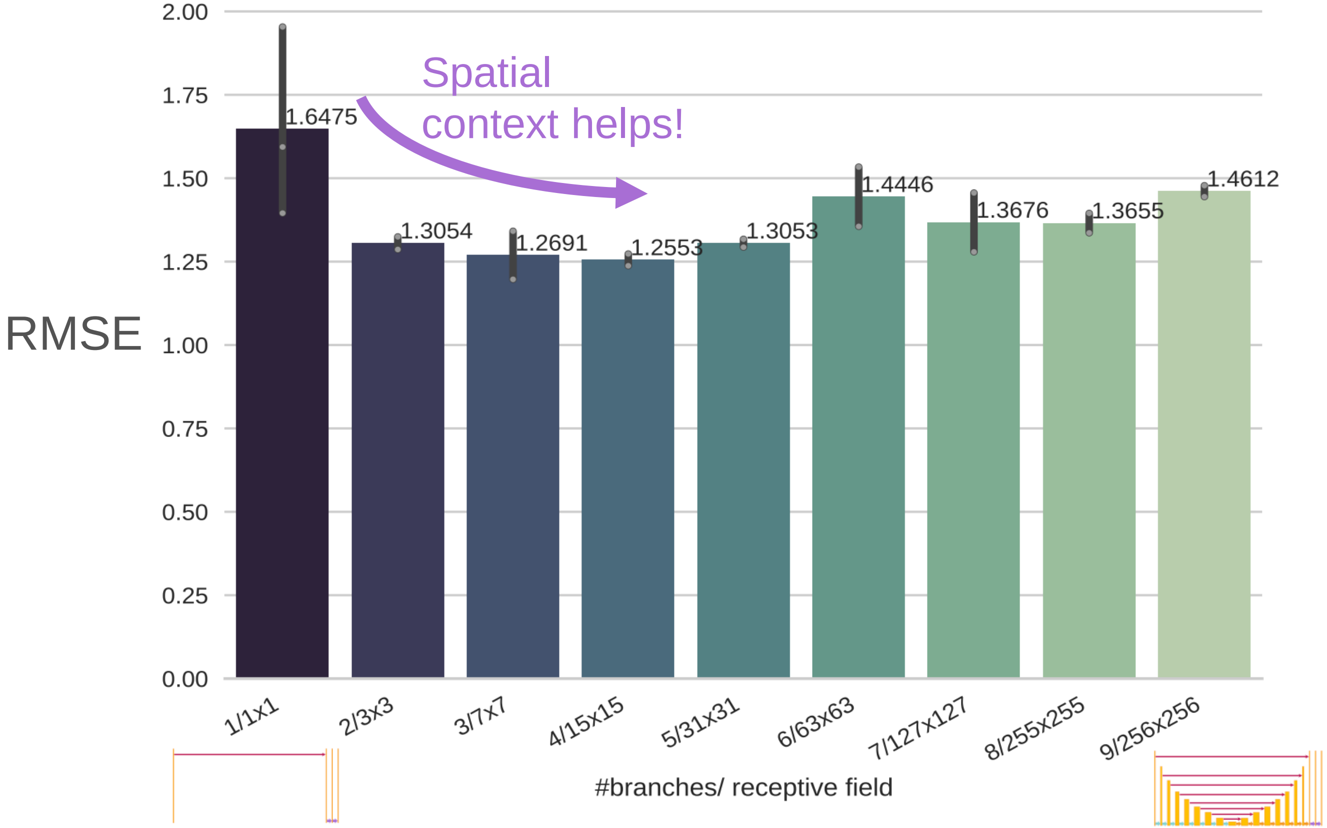
Results

Does spatial information really help?



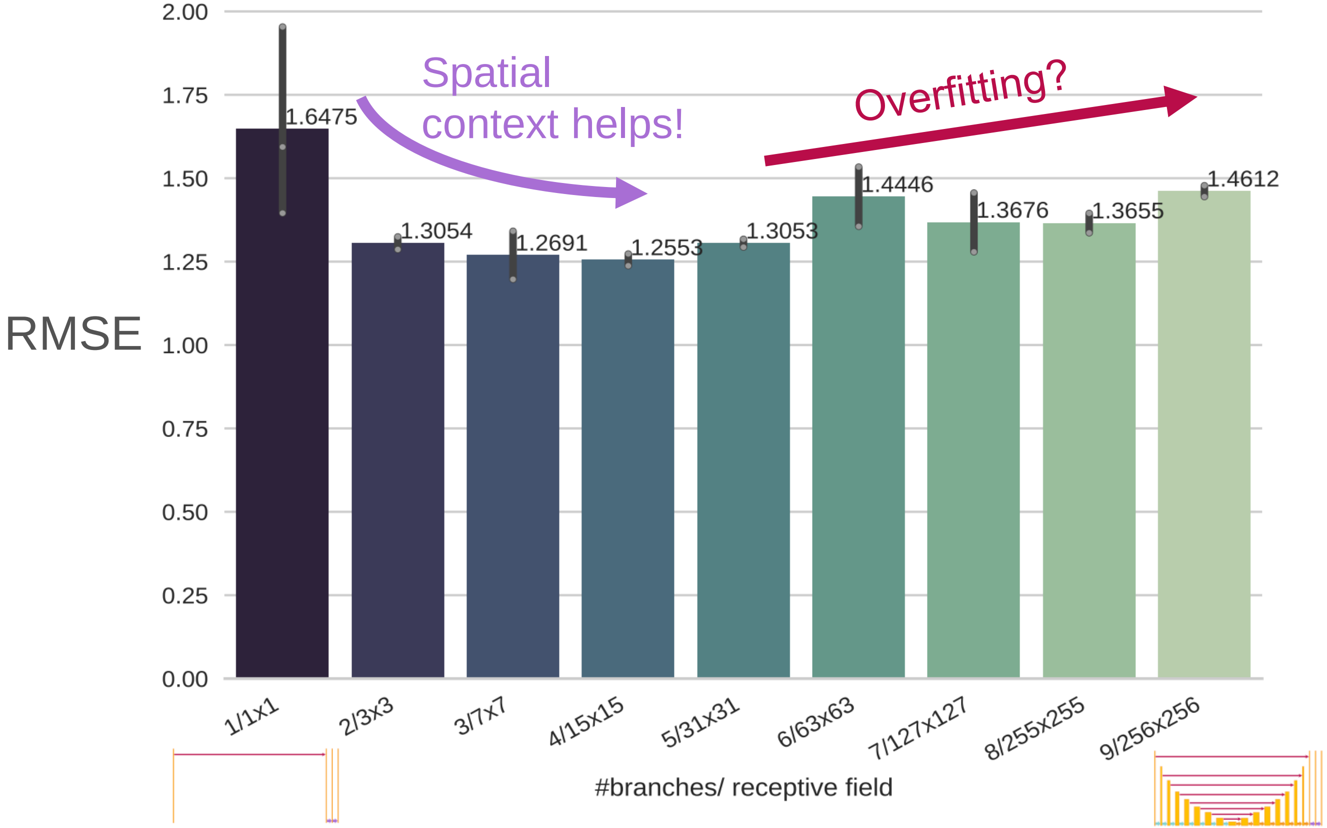
Results

Does spatial information really help?



Results

Does spatial information really help?



Takeaways

- CNNs/GANs applied to spectral image reconstruction for the 1st time
- State of the art over ICVL dataset
- Spatial context information does help*

Publication **A. Alvarez-Gila**, J. van de Weijer, and E. Garrote, "*Adversarial Networks for Spatial Context-Aware Spectral Image Reconstruction from RGB,*" ICCVW 2017

Publication B. Arad, O. Ben-Shahar, R. Timofte, L. Van Gool, L. Zhang, M.-H. Yang, Z. Xiong, C. Chen, Z. Shi, D. Liu, F. Wu, C. Lanaras, S. Galliani, K. Schindler, T. Stiebel, S. Koppers, P. Seltsam, R. Zhou, M. El Helou, F. Lahoud, M. Shahpaski, K. Zheng, L. Gao, B. Zhang, X. Cui, H. Yu, Y. B. Can, **A. Alvarez-Gila**, J. van de Weijer, E. Garrote, A. Galdran, M. Sharma, S. Koundinya, A. Upadhyay, R. Manekar, R. Mukhopadhyay, H. Sharma, S. Chaudhury, K. Nagasubramanian, S. Ghosal, A. K. Singh, A. Singh, B. Ganapathysubramanian, and S. Sarkar, "*NTIRE 2018 Challenge on Spectral Reconstruction from RGB Images,*" CVPRW 2018

https://github.com/aitorshuffle/ntire2018_adv_rgb2hs

Self-supervised learning for image-to-image translation in the small data regime

Context

Steelmaking with an Electric Arc Furnace

FluidMediaStudios, "*Electric Arc Furnace Operations*," 2019. https://www.youtube.com/watch?v=HKQ2GaXFI3w

Context

Steelmaking with an Electric Arc Furnace

- Holy Grail of EAF-based steelmaking:

Remote, online estimation of slag composition

e.g.

SiO_2 (%)	FeO (%)	Al_2O_3 (%)	CaO (%)	MgO (%)
25.24	0.23	6.15	60.02	3.49

- State of the affairs:
  - Manual temperature mesasurement through thermocouple
  - Offline chemical analysis of cooled (solid) preprocessed slag sample (XRF spectrometry)

Problem statement

- Holy Grail of EAF-based steelmaking:

Remote, online estimation of slag ~~composition~~

temperature and **spectral emissivity**

i.e. Temperature-Emissivity Separation (TES)

Problem statement

- Holy Grail of EAF-based steelmaking:

Remote, online estimation of slag ~~composition~~

temperature and **spectral emissivity**

i.e. Temperature-Emissivity Separation (TES)

Useful process variable

Problem statement

- Holy Grail of EAF-based steelmaking:

Remote, online estimation of slag ~~composition~~

temperature and **spectral emissivity**

i.e. Temperature-Emissivity Separation (TES)

Useful process variable

Proxy for composition

Problem statement

- Holy Grail of EAF-based steelmaking:

Remote, online estimation of slag ~~composition~~
temperature and **spectral emissivity**

i.e. Temperature-Emissivity Separation (TES)

Useful process variable

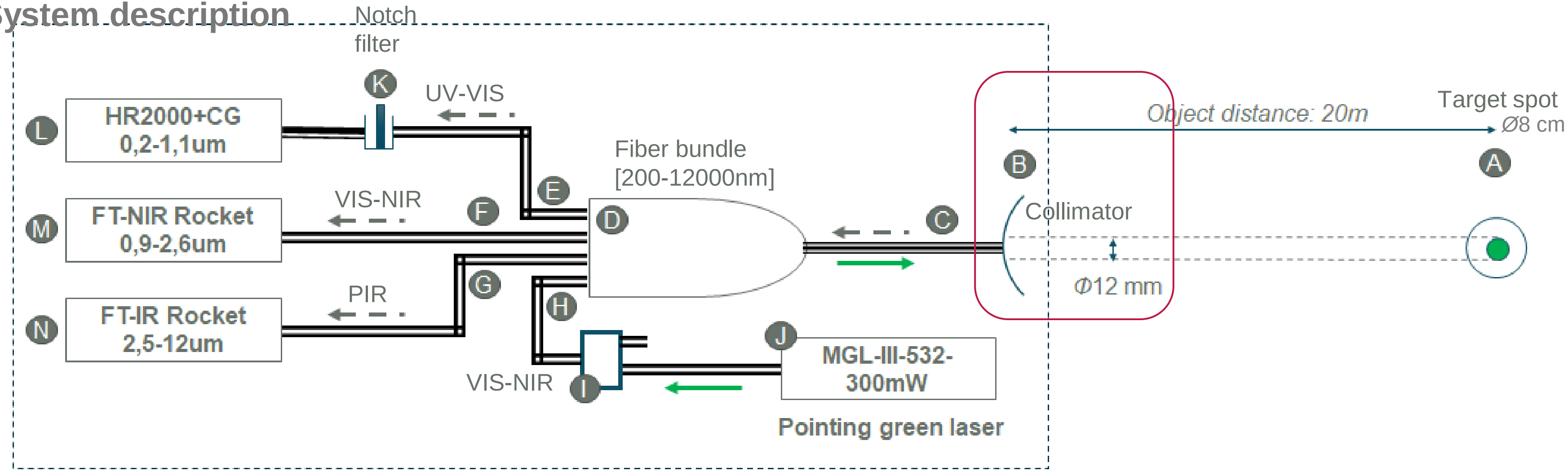
Proxy for composition

- Heavily underconstrained problem
- Previous methods pose strong assumptions:
  - Uniform spectral emissivity [Rego-Barcena2008]
  - Known temperature [Lee2013]
  - Specific temperature/emissivity ranges (e.g. remote sensing) [Barducci2014]

Chapter	Small data	Technique	Prior knowledge/Physics	Synthetic
Ch.4 (TES)	No labels	PP	Radiative Transfer	-

Design of the device

System description



Prototype:

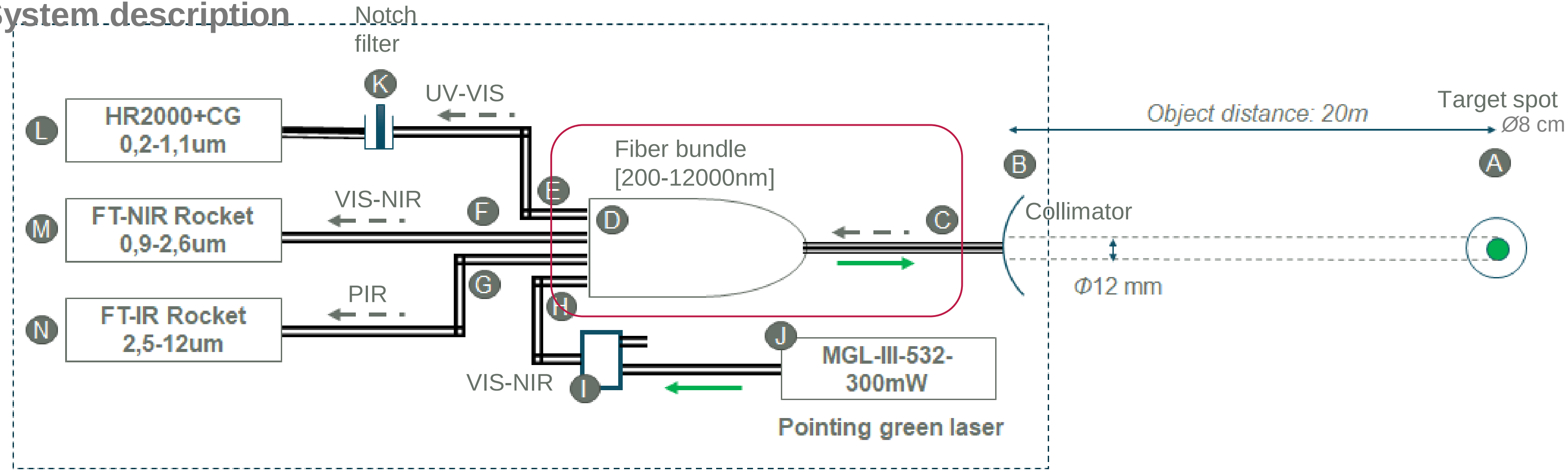
Acquisition case

Opto-mecanical system

Tripod mount

Design of the device

System description

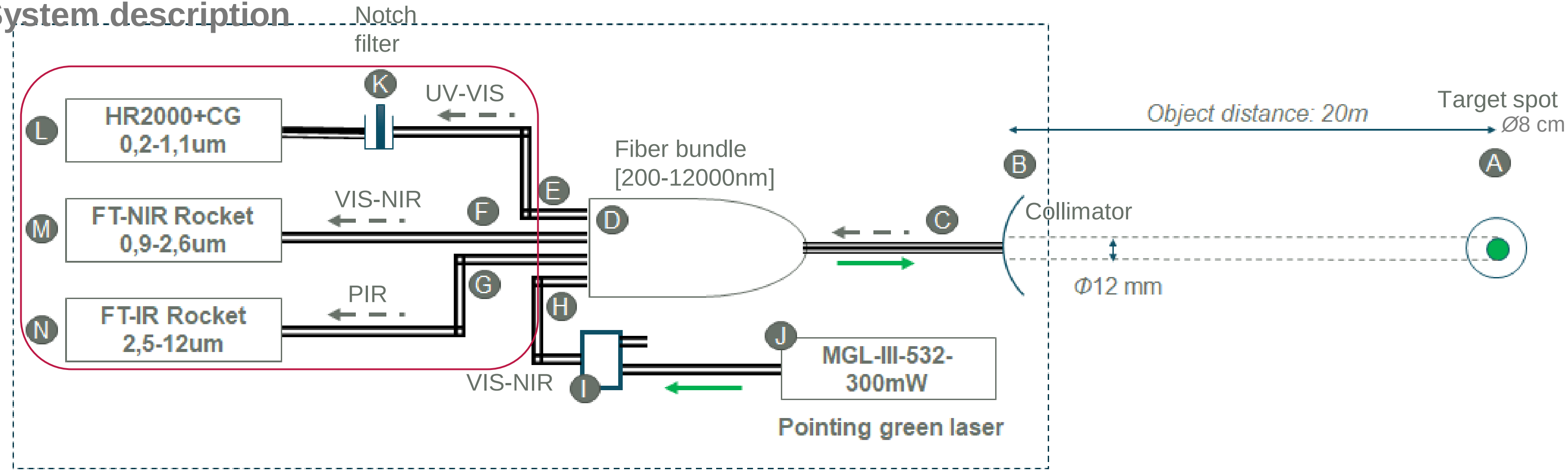




Prototype: Acquisition case Opto-mecanical system Tripod mount

Design of the device

System description



Prototype: Acquisition case | Opto-mecanical system | Tripod mount

System calibration

Offline calibration. Laboratory blackbody (BB) furnace [500-1500ºC]

- Corrects sensor non-linearities. Maps the collected counts by the spectrometer at each λ_i into the theoretical blackbody radiance at T_j.

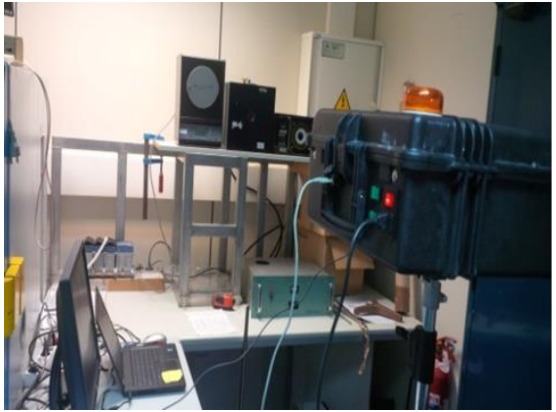
$$\underset{\boxed{W_{\lambda_i}}}{\arg\min} \sqrt{\sum_{T_j=T_0}^{T_n} [\boxed{L_{bb_{\lambda_i}}}(T_j) - \boxed{f_{calib}}(\boxed{C_{\lambda_i}(T_j)}, \boxed{W_{\lambda_i}})]^2}$$

Theoretical BB radiance at temperature T_j

2nd degree poly

Collected spectrometer **counts** per wavelength λ_i at BB temperature T_j

Per-wavelength polynomial **coefficients**



Offline calibration setup

Field calibration. Portable calibration lamp

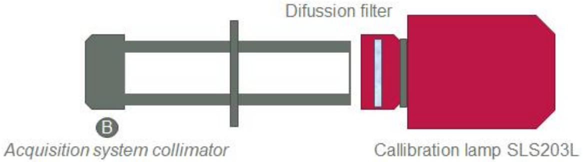
- Accounts for daily mechanical variations (up to 15%)

Linear transmission coefficients

$$K_s = \frac{\sum_{\lambda=\lambda_l}^{\lambda_h} \boxed{L_{s1}(\lambda)}}{\sum_{\lambda=\lambda_l}^{\lambda_h} \boxed{L_{s0}(\lambda)}}, \quad \forall s \in \{1, 2, 3\}$$

Current lamp-induced radiance for spectrometer S

Lamp-induced radiance for spectrometer S after BB calibration



Calibration lamp [500-900nm], 1500K

4-TES

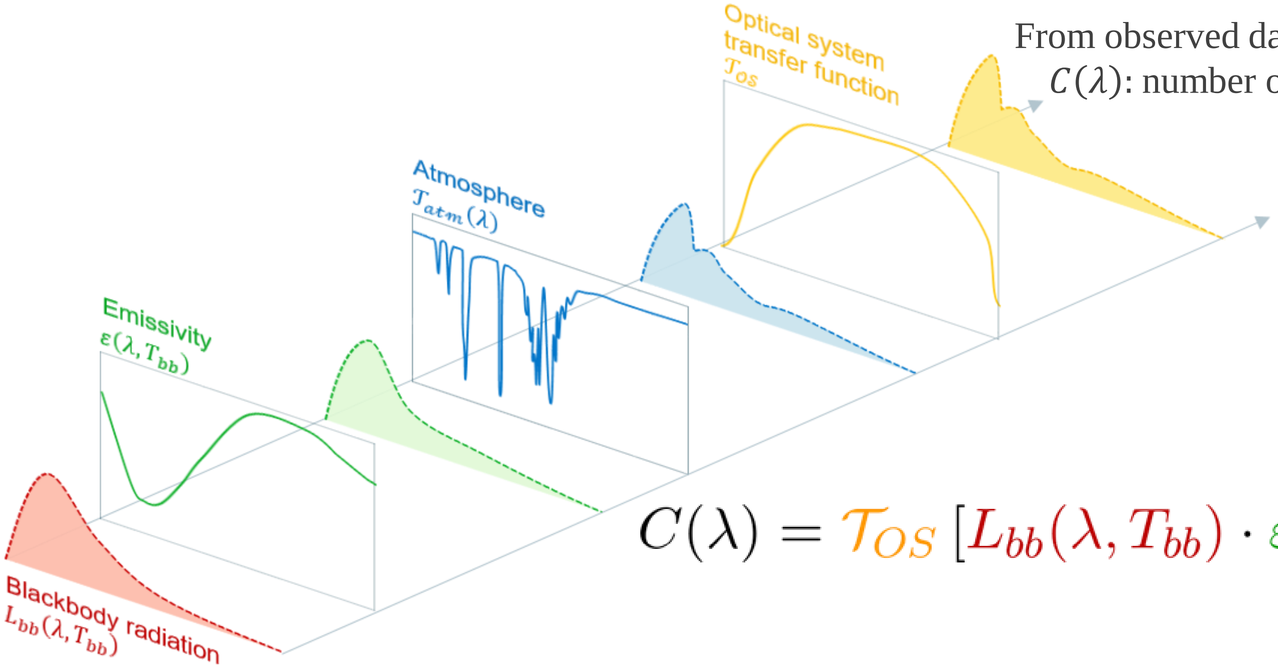
Radiative transfer model

Model formulation

Goal: TES - simultaneous estimation of the temperature $\boldsymbol{T_{bb}}$ and spectral emissivity $\boldsymbol{\varepsilon(\lambda, T_{bb})}$ of the observed hot sample

From observed data:

$C(\lambda)$: number of counts per wavelength

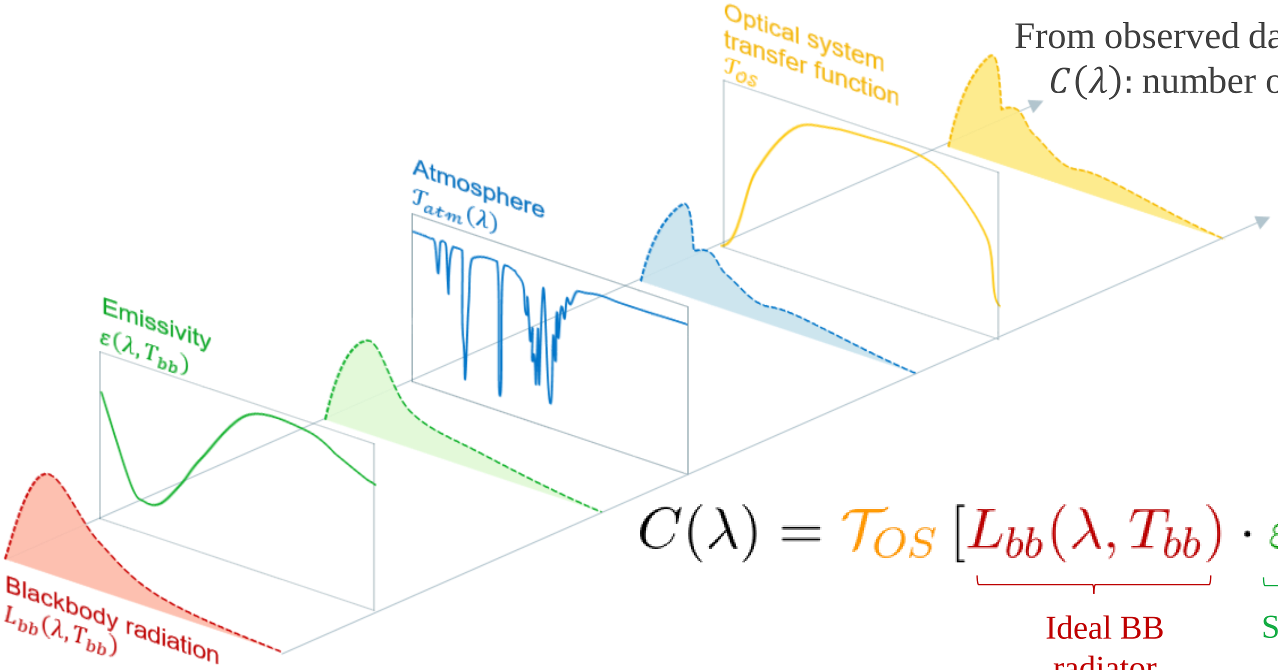


$$C(\lambda) = \mathcal{T}_{OS}\left[L_{bb}(\lambda, T_{bb}) \cdot \varepsilon(\lambda, T_{bb}) \cdot \mathcal{T}_{atm}(\lambda)\right]$$

Radiative transfer model

Model formulation

Goal: TES - simultaneous estimation of the temperature $\boldsymbol{T_{bb}}$ and spectral emissivity $\boldsymbol{\varepsilon(\lambda, T_{bb})}$ of the observed hot sample



From observed data:

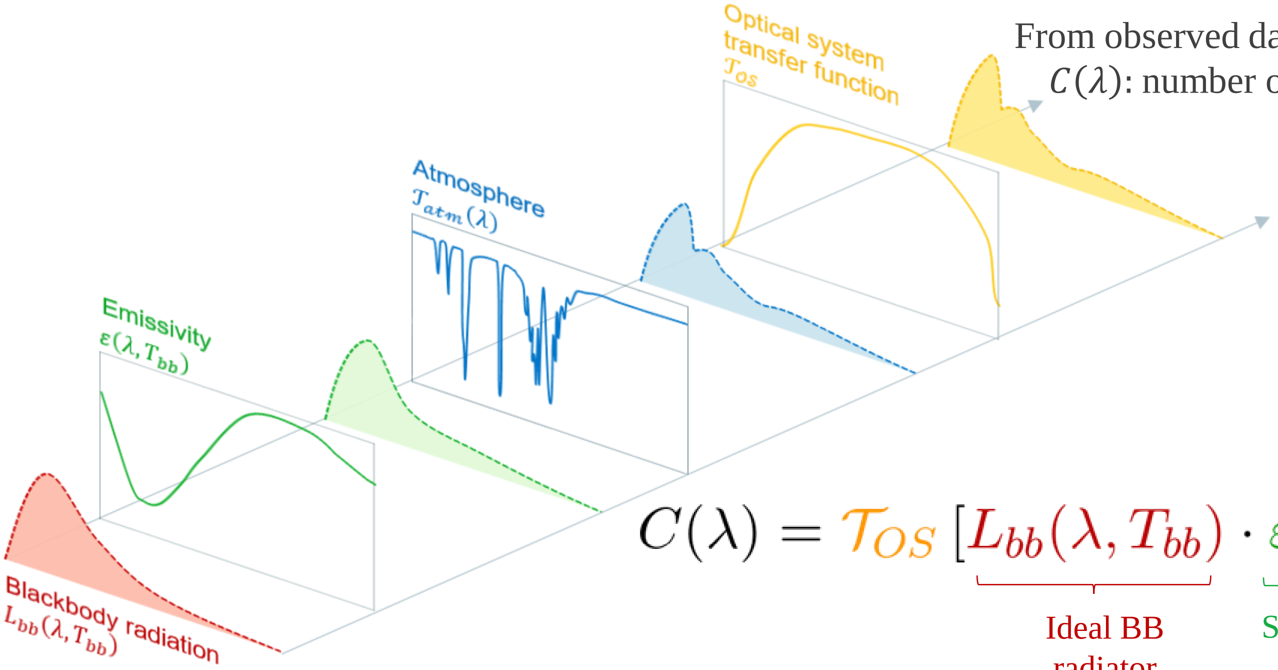
$C(\lambda)$: number of counts per wavelength

$$C(\lambda) = \mathcal{T}_{OS}\left[\underbrace{\underbrace{L_{bb}(\lambda, T_{bb})}_{\text{Ideal BB radiator}} \cdot \underbrace{\varepsilon(\lambda, T_{bb})}_{\text{Sample spectral emissivity}}}_{\text{Selective radiator (emissive sample)}} \cdot \mathcal{T}_{atm}(\lambda)\right]$$

Radiative transfer model

Model formulation

Goal: TES - simultaneous estimation of the temperature $\boldsymbol{T_{bb}}$ and spectral emissivity $\boldsymbol{\varepsilon(\lambda, T_{bb})}$ of the observed hot sample



From observed data:

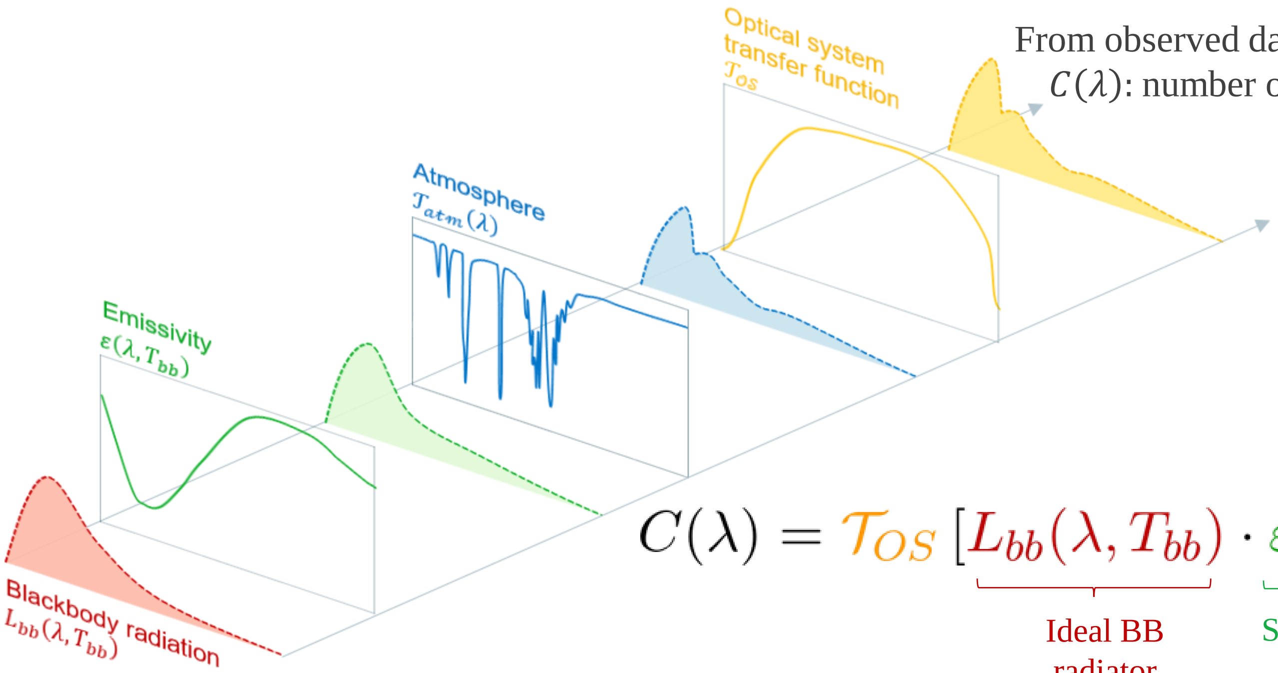
$C(\lambda)$: number of counts per wavelength

$$C(\lambda) = \mathcal{T}_{OS}\,\underbrace{\Big[\underbrace{L_{bb}(\lambda, T_{bb})}_{\text{Ideal BB radiator}} \cdot \underbrace{\varepsilon(\lambda, T_{bb})}_{\text{Sample spectral emissivity}}}_{\text{Selective radiator (emissive sample)}} \cdot \underbrace{\mathcal{T}_{atm}(\lambda)}_{\text{Atmospheric transmittance}}\Big]$$

Radiative transfer model

Model formulation

Goal: TES - simultaneous estimation of the temperature $\boldsymbol{T_{bb}}$ and spectral emissivity $\boldsymbol{\varepsilon(\lambda, T_{bb})}$ of the observed hot sample



From observed data:

$C(\lambda)$: number of counts per wavelength

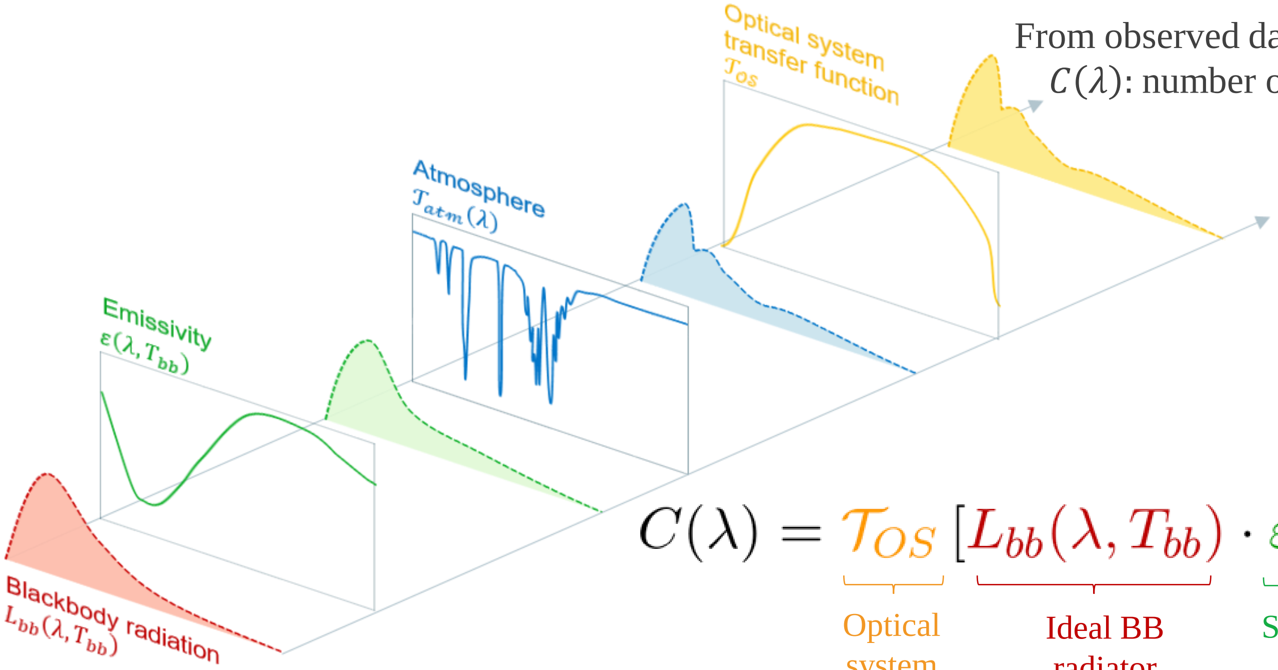
$$C(\lambda) = \mathcal{T}_{OS} \left[\underbrace{\underbrace{L_{bb}(\lambda, T_{bb})}_{\text{Ideal BB radiator}} \cdot \underbrace{\varepsilon(\lambda, T_{bb})}_{\text{Sample spectral emissivity}}}_{\text{Selective radiator (emissive sample)}} \cdot \underbrace{\mathcal{T}_{atm}(\lambda)}_{\text{Atmospheric transmittance}} \right]$$

Observed radiance at entry of capturing device

Radiative transfer model

Model formulation

Goal: TES - simultaneous estimation of the temperature $\boldsymbol{T_{bb}}$ and spectral emissivity $\boldsymbol{\varepsilon(\lambda, T_{bb})}$ of the observed hot sample



From observed data:

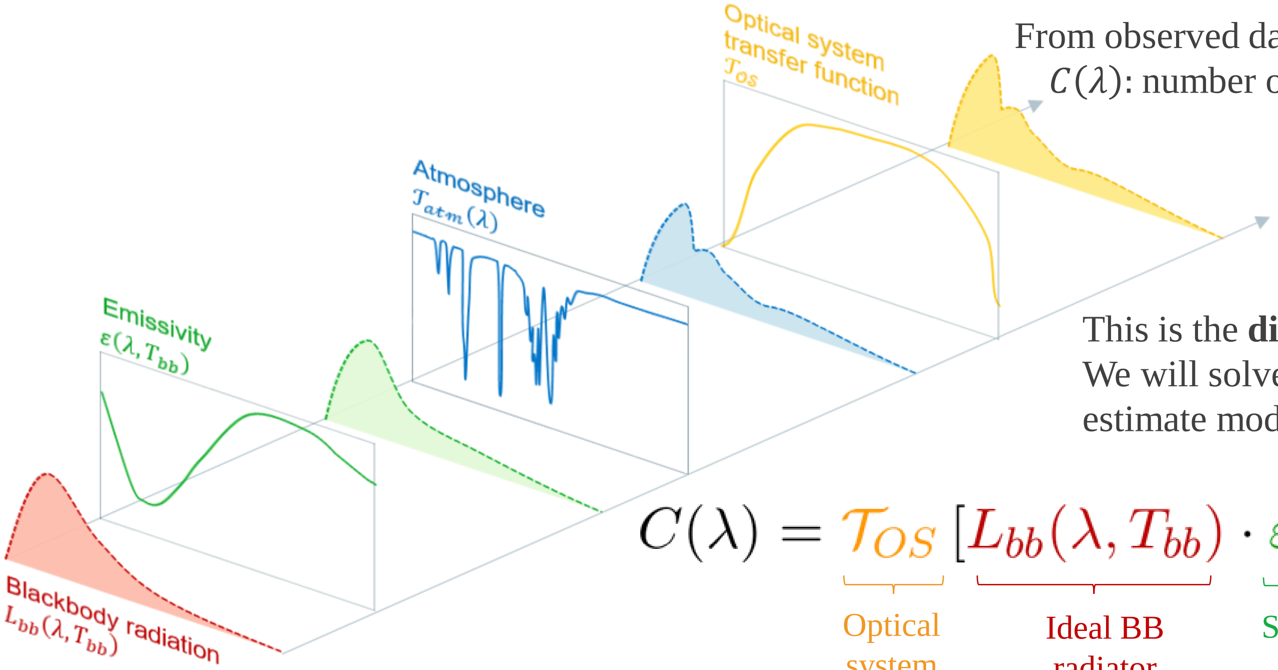
$C(\lambda)$: number of counts per wavelength

$$C(\lambda) = \underbrace{\mathcal{T}_{OS}}_{\text{Optical system}} \underbrace{\Big[\underbrace{\underbrace{L_{bb}(\lambda, T_{bb})}_{\text{Ideal BB radiator}} \cdot \underbrace{\varepsilon(\lambda, T_{bb})}_{\text{Sample spectral emissivity}}}_{\text{Selective radiator (emissive sample)}} \cdot \underbrace{\mathcal{T}_{atm}(\lambda)}_{\text{Atmospheric transmittance}}\Big]}_{\text{Observed radiance at entry of capturing device}}$$

Radiative transfer model

Model formulation

Goal: TES - simultaneous estimation of the temperature $\boldsymbol{T_{bb}}$ and spectral emissivity $\boldsymbol{\varepsilon(\lambda, T_{bb})}$ of the observed hot sample



From observed data:

$C(\lambda)$: number of counts per wavelength

This is the **direct** model.
We will solve the **inverse** problem:
estimate model parameters from noisy observations

$$\underbrace{\underbrace{C(\lambda) = \underbrace{\mathcal{T}_{OS}}_{\text{Optical system}} [\underbrace{\underbrace{L_{bb}(\lambda, T_{bb})}_{\text{Ideal BB radiator}} \cdot \underbrace{\varepsilon(\lambda, T_{bb})}_{\text{Sample spectral emissivity}}}_{\text{Selective radiator (emissive sample)}} \cdot \underbrace{\mathcal{T}_{atm}(\lambda)}_{\text{Atmospheric transmittance}}]}_{\text{Observed radiance at entry of capturing device}}}$$

Radiative transfer model

Model formulation: Ideal blackbody radiator $L_{bb}(\lambda, T_{bb})$

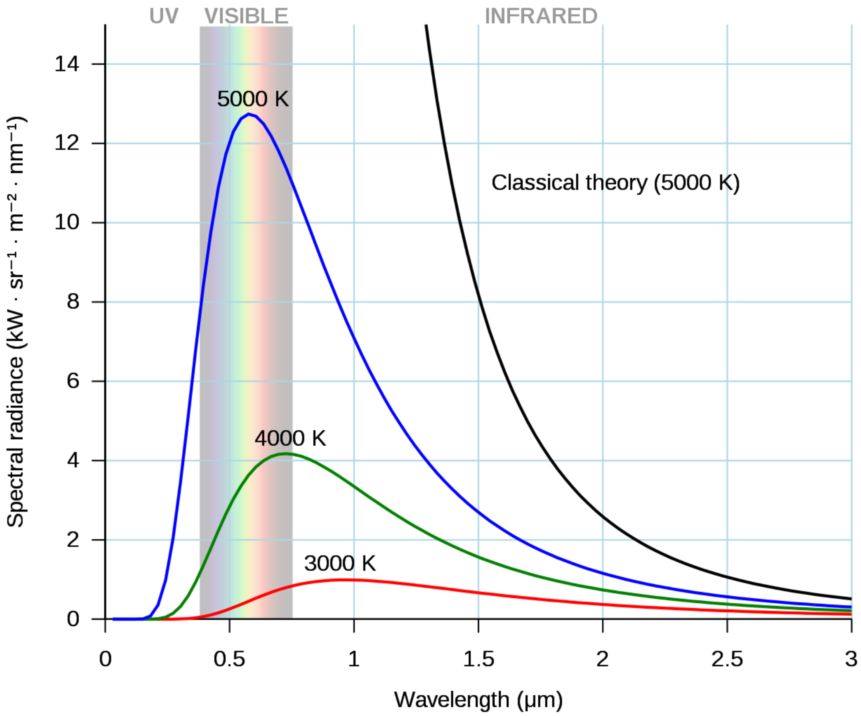
Planck's law of black-body radiation: Spectral radiance of a BB at temperature T_{bb}:

Planck constant

Speed of light

$$L_{bb}(\lambda, T_{bb}) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T_{bb}}} - 1}$$

Boltzmann constant



Radiative transfer model

Model formulation: Ideal blackbody radiator $L_{bb}(\lambda, T_{bb})$

Planck's law of black-body radiation: Spectral radiance of a BB at temperature T_{bb}:

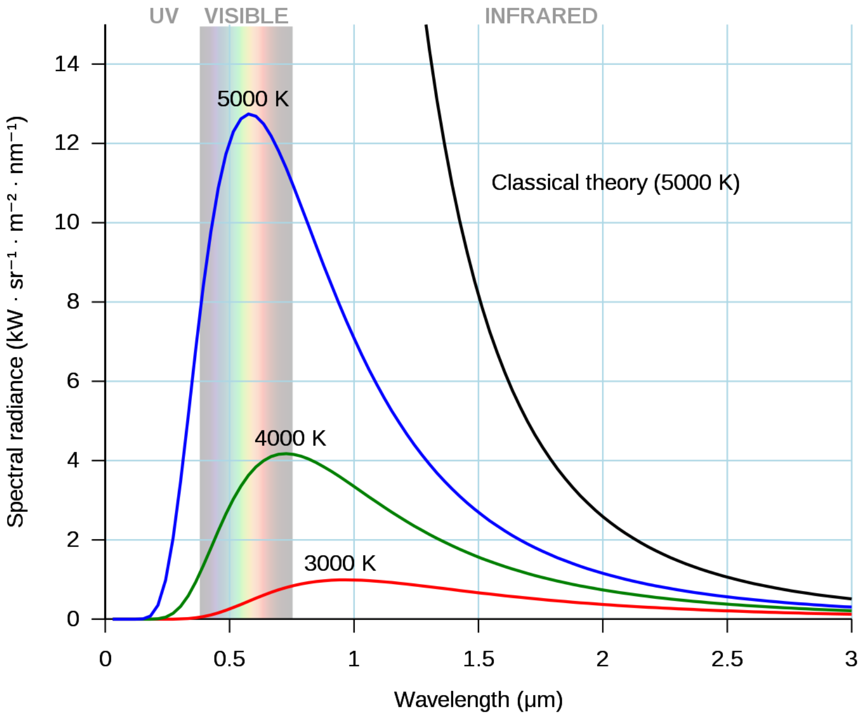
Planck constant

Speed of light

$$L_{bb}(\lambda, T_{bb}) = \frac{2hc^2}{\boxed{\lambda^5}} \frac{1}{\boxed{e}^{\frac{hc}{\boxed{\lambda} k_B \boxed{T_{bb}}}} - 1}$$

Boltzmann constant

A correct estimation is critical!



Radiative transfer model

Model formulation: Ideal blackbody radiator $L_{bb}(\lambda, T_{bb})$

Planck's law of black-body radiation: Spectral radiance of a BB at temperature T_{bb}:

Planck constant

Speed of light

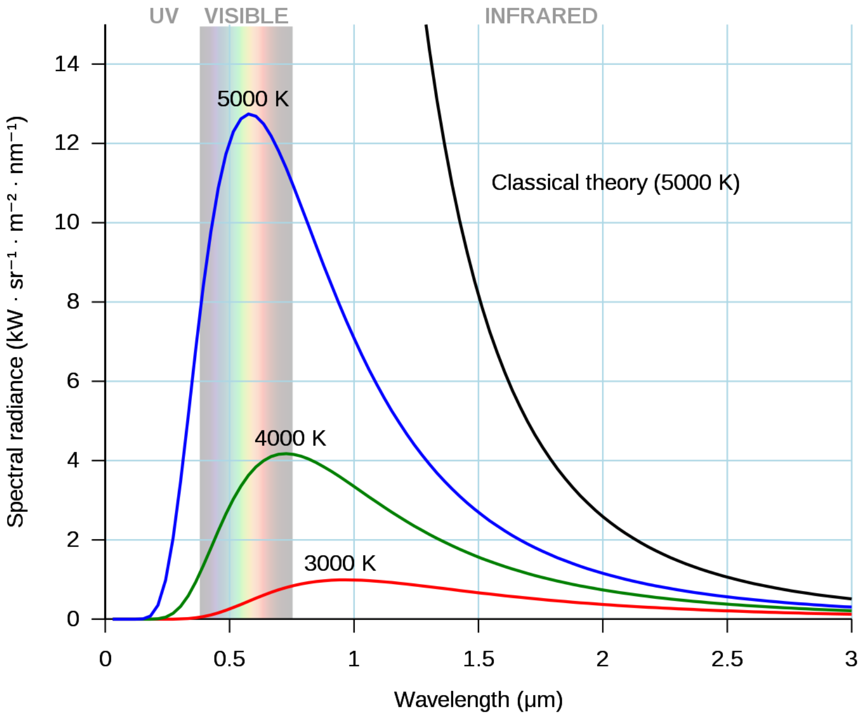
$$L_{bb}(\lambda, T_{bb}) = \frac{2hc^2}{\boxed{\lambda^5}} \frac{1}{\boxed{e}^{\frac{hc}{\boxed{\lambda} k_B \boxed{T_{bb}}}} - 1}$$

Boltzmann constant

A correct estimation is critical!

List of stochastic variables:

- T_{bb}

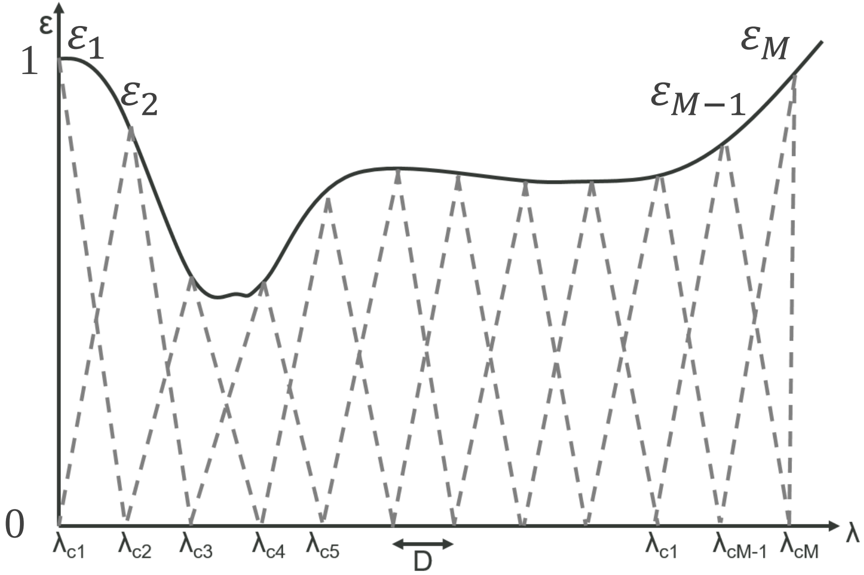


4-TES

Radiative transfer model

Model formulation: Spectral emissivity $\varepsilon(\lambda, T_{bb})$ of the radiative source

- Modeled as M=10 probabilistic variables ε_k with $k = 1 \dots M$
- M associated fuzzy sets with triangular membership
- Spectral emissivity at any λ_i as weighted value over ε_k



List of stochastic variables:

- T_{bb}
- $\varepsilon_k, k = 1 \dots M$

Radiative transfer model

Model formulation: Atmospheric transmitance $T_{atm}(\lambda)$

$$\mathcal{T}_{atm}(\lambda) = e^{-d\cdot\gamma_{atm}} = e^{-d\cdot\sum\limits_{a} -x_a\cdot\gamma_a} = e^{-d\cdot(x_{CO_2}\cdot\gamma_{CO_2}+x_{H_2O}\cdot\gamma_{H_2O})}$$

List of stochastic variables:

- T_{bb}
- $\varepsilon_k, k = 1...M$
- x_{CO_2}, x_{H_2O}

4-TES

Radiative transfer model

Model formulation: Atmospheric transmitance $T_{atm}(\lambda)$

Lambert-Beer

Unitary absorption coefficient

$$\mathcal{T}_{atm}(\lambda) = e^{-d\cdot\gamma_{atm}} = e^{-d\cdot\sum_a -x_a\cdot\gamma_a} = e^{-d\cdot(x_{CO_2}\cdot\gamma_{CO_2}+x_{H_2O}\cdot\gamma_{H_2O})}$$

Distance (set manually)

Absorbant a

Molar concentration

Focus on CO_2 and H_2O

List of stochastic variables:

- T_{bb}
- $\varepsilon_k, k = 1...M$
- x_{CO_2}, x_{H_2O}

Radiative transfer model

Model formulation: Atmospheric transmitance $T_{atm}(\lambda)$

Lambert-Beer

Unitary absorption coefficient

$$\mathcal{T}_{atm}(\lambda) = e^{-d\cdot\gamma_{atm}} = e^{-d\cdot\sum_a -x_a\cdot\gamma_a} = e^{-d\cdot(x_{CO_2}\cdot\gamma_{CO_2}+x_{H_2O}\cdot\gamma_{H_2O})}$$

Distance (set manually)

Absorbant a

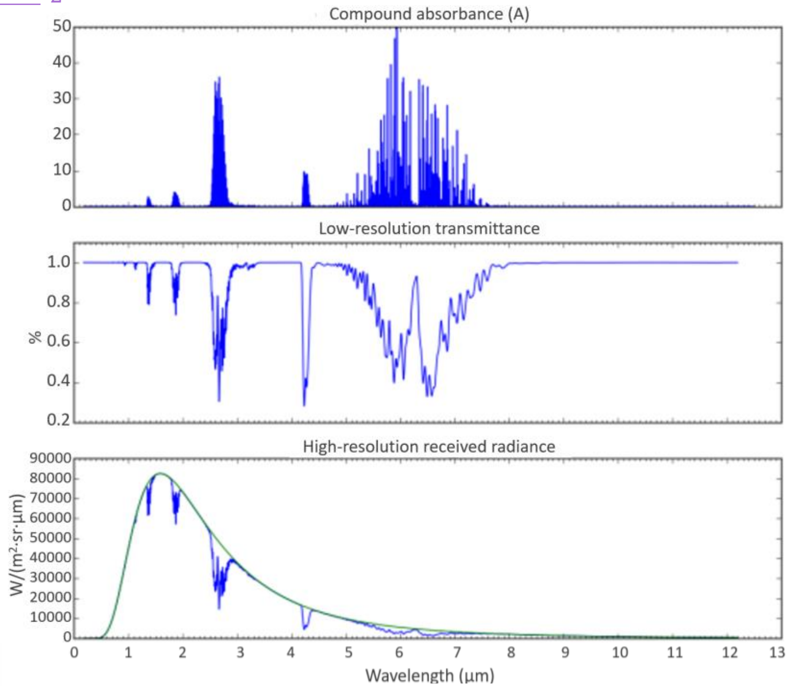
Molar concentration

Focus on CO_2 and H_2O

Combined absorbance
Line-by-line cross-section information.
Modeled using HITRAN2016's API (HAPI)

Equivalent (low-res) transmittance $T_{atm}(\lambda)$
Convolved with spectrometers' slit functions

BB radiance and
atmosphere-filtered radiance



Radiance of ideal BB ($T_{bb} = 1550ºC$), filtered by simulated $T_{atm}(\lambda)$, d = $1.5m, T = 27ºC$. Typical concentrations.

List of stochastic variables:

- T_{bb}
- $\varepsilon_k, k = 1 \ldots M$
- x_{CO_2}, x_{H_2O}

Radiative transfer model

Model formulation: Final

Variations over precomputed K_s.

$$L_{expected}(\lambda) = \mathcal{T}_{OS}^{-1}\underbrace{\left[\mathcal{T}_{OS}\left(k_s \cdot L_{bb}(\lambda, T_{bb}) \cdot \varepsilon(\lambda, T_{bb}) \cdot e^{-\gamma_{atm}(\lambda)\cdot d}\right)\right]}_{C(\lambda)}$$

Radiative transfer model

Model formulation: Final

$$L_{expected}(\lambda) = \mathcal{T}_{OS}^{-1}\underbrace{\left[\overbrace{\mathcal{T}_{OS}}^{\text{Variations over precomputed } K_s.}\left(k_s \cdot L_{bb}(\lambda, T_{bb}) \cdot \varepsilon(\lambda, T_{bb}) \cdot e^{-\gamma_{atm}(\lambda)\cdot d}\right)\right]}_{C(\lambda)}$$

List of stochastic variables:

- T_{bb}
- x_{CO_2}, x_{H_2O}
- $\varepsilon_k, k = 1...M$
- $k_s, s = 1, 2, 3$

$$\theta = \{T_{bb}, \sigma, x_{CO_2}, x_{H_2O}, k_1, k_2, k_3, \varepsilon_k\} \; \forall k \in [1 \ldots M].$$

We will solve this through Bayesian inference
(Probabilistic Programming with PyMC3)

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)} \qquad \boldsymbol{Bayes'\ formula}$$

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$$\boxed{P(\theta|x)} = \frac{P(x|\theta)P(\theta)}{P(x)}$$

$\boldsymbol{Bayes' \; formula}$

$\boldsymbol{P(\theta|x)}$ [Posterior probability distribution]

Our quantity of interest:

$P(\theta|x) = P\left(T_{bb}, \sigma, x_{CO_2}, x_{H_2O}, k_1, k_2, k_3, \varepsilon_1, \ldots \varepsilon_M \middle| L_{obs}(\lambda)\right)$

What are the values of these parameters that best explain the measured radiance data ($L_{obs}(\lambda)$)?

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$\boldsymbol{P(\theta)}$ **[Prior]**
Incorporates our prior knowledge over the parameter values.

$$\boxed{P(\theta|x)} = \frac{P(x|\theta)\boxed{P(\theta)}}{P(x)} \qquad \boldsymbol{Bayes'\ formula}$$

$\boldsymbol{P(\theta|x)}$ **[Posterior probability distribution]**
Our quantity of interest:

$P(\theta|x) = P\left(T_{bb}, \sigma, x_{CO_2}, x_{H_2O}, k_1, k_2, k_3, \varepsilon_1, \ldots \varepsilon_M \middle| L_{obs}(\lambda)\right)$
What are the values of these parameters that best explain the measured radiance data ($L_{obs}(\lambda)$)?

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$\boldsymbol{P(\theta)}$ **[Prior]**
Incorporates our prior knowledge over the parameter values.

$\boldsymbol{P(x|\theta)}$ **[Likelihood]**
Describes how the data was generated.
Our forward model

$$\boxed{P(\theta|x)} = \frac{\boxed{P(x|\theta)}\boxed{P(\theta)}}{P(x)}$$

$\boldsymbol{Bayes'\ formula}$

$\boldsymbol{P(\theta|x)}$ **[Posterior probability distribution]**
Our quantity of interest:

$P(\theta|x) = P\left(T_{bb}, \sigma, x_{CO_2}, x_{H_2O}, k_1, k_2, k_3, \varepsilon_1, \dots \varepsilon_M \middle| L_{obs}(\lambda)\right)$
What are the values of these parameters that best explain the measured radiance data ($L_{obs}(\lambda)$)?

4-TES

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$\boldsymbol{P(x|\theta)}$ **[Likelihood]**
Describes how the data was generated.
Our forward model

$\boldsymbol{P(\theta)}$ **[Prior]**
Incorporates our prior knowledge over the parameter values.

$$\boxed{P(\theta|x)} = \frac{\boxed{P(x|\theta)}\boxed{P(\theta)}}{\boxed{P(x)}}$$

$\boldsymbol{Bayes'\ formula}$

$\boldsymbol{P(x)}$ **[Evidence]**
$P(x) = \int_\theta P(x,\theta)d\theta$ (No closed-form solution)

$\boldsymbol{P(\theta|x)}$ **[Posterior probability distribution]**
Our quantity of interest:

$P(\theta|x) = P\left(T_{bb}, \sigma, x_{CO_2}, x_{H_2O}, k_1, k_2, k_3, \varepsilon_1, \ldots \varepsilon_M \middle| L_{obs}(\lambda)\right)$
What are the values of these parameters that best explain the measured radiance data ($L_{obs}(\lambda)$)?

4-TES

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$\boldsymbol{P(x|\theta)}$ **[Likelihood]**
Describes how the data was generated.
Our forward model

$\boldsymbol{P(\theta)}$ **[Prior]**
Incorporates our prior knowledge over the parameter values.

$$\boxed{P(\theta|x)} = \frac{\boxed{P(x|\theta)}\boxed{P(\theta)}}{\boxed{P(x)}}$$

$\boldsymbol{Bayes' \ formula}$

$\boldsymbol{P(x)}$ **[Evidence]**
$P(x) = \int_\theta P(x,\theta) d\theta$ (No closed-form solution)

$\boldsymbol{P(\theta|x)}$ **[Posterior probability distribution]**
Our quantity of interest:

$P(\theta|x) = P\left(T_{bb}, \sigma, x_{CO_2}, x_{H_2O}, k_1, k_2, k_3, \varepsilon_1, \ldots \varepsilon_M \middle| L_{obs}(\lambda)\right)$
What are the values of these parameters that best explain the measured radiance data ($L_{obs}(\lambda)$)?

Build a Markov Chain generating samples, with a distribution that matches the posterior:

$$P(\theta|x) \propto P(x|\theta)P(\theta)$$

Simultaneous estimation of every parameter with a measure of its **uncertainty**

We only need likelihood and priors!

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$$L_{expected}(\lambda) = \mathcal{T}_{OS}^{-1}\left[\mathcal{T}_{OS}\left(k_s \cdot L_{bb}(\lambda, T_{bb}) \cdot \varepsilon(\lambda, T_{bb}) \cdot e^{-\gamma_{atm}(\lambda)\cdot d}\right)\right]$$

Parameter	Name	Prior distribution
T_{bb}	Sample temperature	$\sim \mathcal{U}(min = 400°C, max = 1500°C)$
x_{CO_2}	Molar concentration of CO_2	$\sim \mathcal{N}(\mu = 450ppm, \sigma^2 = 50^2)$
x_{H_2O}	Molar concentration of H_2O	$\sim \mathcal{N}(\mu = 36000ppm, \sigma^2 = 500^2)$
ε_k	Spectral emissivity anchor value	$\sim \mathcal{U}(min = 0.0, max = 1.0)$
k_s	Spectrometer-wise misalignment proportionality constant	$\sim \mathcal{N}(\mu = [\text{closed form } K_s], \sigma^2 = 0.001^2)$
L_{obs}	Observed radiance	$\sim \mathcal{N}(\mu = [L_{expected}(\lambda)], \sigma^2)$
σ	Standard deviation of the $\mathcal{N}$ modeling L_{obs}	$\sim HalfCauchy(\beta = 10)$

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$$\boxed{L_{expected}(\lambda) = \mathcal{T}_{OS}^{-1}\left[\mathcal{T}_{OS}\left(k_s \cdot L_{bb}(\lambda, T_{bb}) \cdot \varepsilon(\lambda, T_{bb}) \cdot e^{-\gamma_{atm}(\lambda)\cdot d}\right)\right]}$$

Parameter	Name	Prior distribution
T_{bb}	Sample temperature	$\sim \mathcal{U}(min = 400°C, max = 1500°C)$
x_{CO_2}	Molar concentration of CO_2	$\sim \mathcal{N}(\mu = 450ppm, \sigma^2 = 50^2)$
x_{H_2O}	Molar concentration of H_2O	$\sim \mathcal{N}(\mu = 36000ppm, \sigma^2 = 500^2)$
ε_k	Spectral emissivity anchor value	$\sim \mathcal{U}(min = 0.0, max = 1.0)$
k_s	Spectrometer-wise misalignment proportionality constant	$\sim \mathcal{N}(\mu = [\text{closed form } K_s], \sigma^2 = 0.001^2)$
L_{obs}	Observed radiance	$\sim \mathcal{N}(\mu = [L_{expected}(\lambda)], \sigma^2)$
σ	Standard deviation of the $\mathcal{N}$ modeling L_{obs}	$\sim HalfCauchy(\beta = 10)$

$L_{bb}(\lambda, T_{bb})$

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$$L_{expected}(\lambda) = \mathcal{T}_{OS}^{-1}\left[\mathcal{T}_{OS}\left(k_s \cdot L_{bb}(\lambda, T_{bb}) \cdot \varepsilon(\lambda, T_{bb}) \cdot e^{-\gamma_{atm}(\lambda)\cdot d}\right)\right]$$

Parameter	Name	Prior distribution
T_{bb}	Sample temperature	$\sim \mathcal{U}(min = 400°C, max = 1500°C)$
x_{CO_2}	Molar concentration of CO_2	$\sim \mathcal{N}(\mu = 450ppm, \sigma^2 = 50^2)$
x_{H_2O}	Molar concentration of H_2O	$\sim \mathcal{N}(\mu = 36000ppm, \sigma^2 = 500^2)$
ε_k	Spectral emissivity anchor value	$\sim \mathcal{U}(min = 0.0, max = 1.0)$
k_s	Spectrometer-wise misalignment proportionality constant	$\sim \mathcal{N}(\mu = [\text{closed form } K_s], \sigma^2 = 0.001^2)$
L_{obs}	Observed radiance	$\sim \mathcal{N}(\mu = [L_{expected}(\lambda)], \sigma^2)$
σ	Standard deviation of the $\mathcal{N}$ modeling L_{obs}	$\sim HalfCauchy(\beta = 10)$

$L_{bb}(\lambda, T_{bb})$

$T_{atm}(\lambda)$

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$$\boxed{L_{expected}(\lambda) = \mathcal{T}_{OS}^{-1}\left[\mathcal{T}_{OS}\left(k_s \cdot L_{bb}(\lambda, T_{bb}) \cdot \varepsilon(\lambda, T_{bb}) \cdot e^{-\gamma_{atm}(\lambda)\cdot d}\right)\right]}$$

Parameter	Name	Prior distribution
T_{bb}	Sample temperature	$\sim \mathcal{U}(min = 400°C, max = 1500°C)$
x_{CO_2}	Molar concentration of CO_2	$\sim \mathcal{N}(\mu = 450ppm, \sigma^2 = 50^2)$
x_{H_2O}	Molar concentration of H_2O	$\sim \mathcal{N}(\mu = 36000ppm, \sigma^2 = 500^2)$
ε_k	Spectral emissivity anchor value	$\sim \mathcal{U}(min = 0.0, max = 1.0)$
k_s	Spectrometer-wise misalignment proportionality constant	$\sim \mathcal{N}(\mu = [\text{closed form } K_s], \sigma^2 = 0.001^2)$
L_{obs}	Observed radiance	$\sim \mathcal{N}(\mu = [L_{expected}(\lambda)], \sigma^2)$
σ	Standard deviation of the $\mathcal{N}$ modeling L_{obs}	$\sim HalfCauchy(\beta = 10)$

$L_{bb}(\lambda, T_{bb})$

$T_{atm}(\lambda)$

$\varepsilon(\lambda, T_{bb})$

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$$L_{expected}(\lambda) = \mathcal{T}_{OS}^{-1}\left[\mathcal{T}_{OS}\left(k_s \cdot L_{bb}(\lambda, T_{bb}) \cdot \varepsilon(\lambda, T_{bb}) \cdot e^{-\gamma_{atm}(\lambda)\cdot d}\right)\right]$$

Parameter	Name	Prior distribution
T_{bb}	Sample temperature	$\sim \mathcal{U}(min = 400°C, max = 1500°C)$
x_{CO_2}	Molar concentration of CO_2	$\sim \mathcal{N}(\mu = 450ppm, \sigma^2 = 50^2)$
x_{H_2O}	Molar concentration of H_2O	$\sim \mathcal{N}(\mu = 36000ppm, \sigma^2 = 500^2)$
ε_k	Spectral emissivity anchor value	$\sim \mathcal{U}(min = 0.0, max = 1.0)$
k_s	Spectrometer-wise misalignment proportionality constant	$\sim \mathcal{N}(\mu = [\text{closed form } K_s], \sigma^2 = 0.001^2)$
L_{obs}	Observed radiance	$\sim \mathcal{N}(\mu = [L_{expected}(\lambda)], \sigma^2)$
σ	Standard deviation of the $\mathcal{N}$ modeling L_{obs}	$\sim HalfCauchy(\beta = 10)$

$L_{bb}(\lambda, T_{bb})$

$T_{atm}(\lambda)$

$\varepsilon(\lambda, T_{bb})$

T_{OS}

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$$\boxed{L_{expected}(\lambda) = \mathcal{T}_{OS}^{-1}\left[\mathcal{T}_{OS}\left(k_s \cdot L_{bb}(\lambda, T_{bb}) \cdot \varepsilon(\lambda, T_{bb}) \cdot e^{-\gamma_{atm}(\lambda)\cdot d}\right)\right]}$$

Parameter	Name	Prior distribution
T_{bb}	Sample temperature	$\sim \mathcal{U}(min = 400°C, max = 1500°C)$
x_{CO_2}	Molar concentration of CO_2	$\sim \mathcal{N}(\mu = 450ppm, \sigma^2 = 50^2)$
x_{H_2O}	Molar concentration of H_2O	$\sim \mathcal{N}(\mu = 36000ppm, \sigma^2 = 500^2)$
ε_k	Spectral emissivity anchor value	$\sim \mathcal{U}(min = 0.0, max = 1.0)$
k_s	Spectrometer-wise misalignment proportionality constant	$\sim \mathcal{N}(\mu = [\text{closed form } K_s], \sigma^2 = 0.001^2)$
L_{obs}	Observed radiance	$\sim \mathcal{N}(\mu = [L_{expected}(\lambda)], \sigma^2)$
σ	Standard deviation of the $\mathcal{N}$ modeling L_{obs}	$\sim HalfCauchy(\beta = 10)$

$L_{bb}(\lambda, T_{bb})$

$T_{atm}(\lambda)$

$\varepsilon(\lambda, T_{bb})$

T_{OS}

Radiative transfer model

Solving the model through Bayesian probabilistic inference

$$\boxed{L_{expected}(\lambda) = \mathcal{T}_{OS}^{-1}\left[\mathcal{T}_{OS}\left(k_s \cdot L_{bb}(\lambda, T_{bb}) \cdot \varepsilon(\lambda, T_{bb}) \cdot e^{-\gamma_{atm}(\lambda)\cdot d}\right)\right]}$$

Parameter	Name	Prior distribution
T_{bb}	Sample temperature	$\sim \mathcal{U}(min = 400°C, max = 1500°C)$
x_{CO_2}	Molar concentration of CO_2	$\sim \mathcal{N}(\mu = 450ppm, \sigma^2 = 50^2)$
x_{H_2O}	Molar concentration of H_2O	$\sim \mathcal{N}(\mu = 36000ppm, \sigma^2 = 500^2)$
ε_k	Spectral emissivity anchor value	$\sim \mathcal{U}(min = 0.0, max = 1.0)$
k_s	Spectrometer-wise misalignment proportionality constant	$\sim \mathcal{N}(\mu = [\text{closed form } K_s], \sigma^2 = 0.001^2)$
L_{obs}	Observed radiance	$\sim \mathcal{N}(\mu = [L_{expected}(\lambda)], \sigma^2)$
σ	Standard deviation of the $\mathcal{N}$ modeling L_{obs}	$\sim HalfCauchy(\beta = 10)$

$L_{bb}(\lambda, T_{bb})$

$T_{atm}(\lambda)$

$\varepsilon(\lambda, T_{bb})$

T_{OS}

Our *likelihood* term: we model the ***error between the observed and expected radiances*** as a **normal** distribution

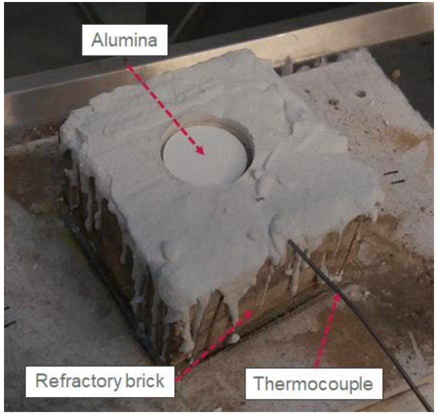
Experimental validation

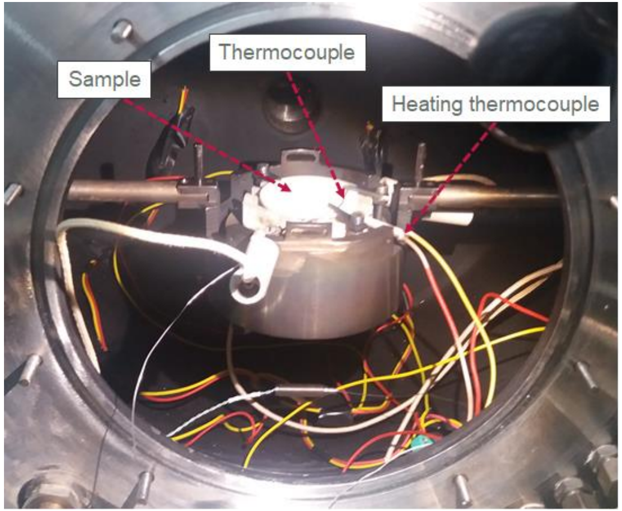
Setup

Testing is difficult!

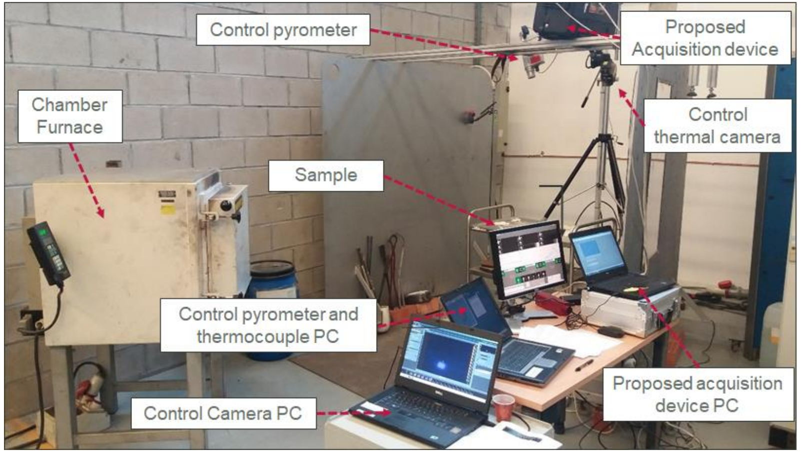
Two solid samples with well-characterized emissivities:

- Alumina (Al_2O_3),
- Boron nitride (BN)





Laboratory equipment (HAIRL emissometer).
Range: 100-860ºC

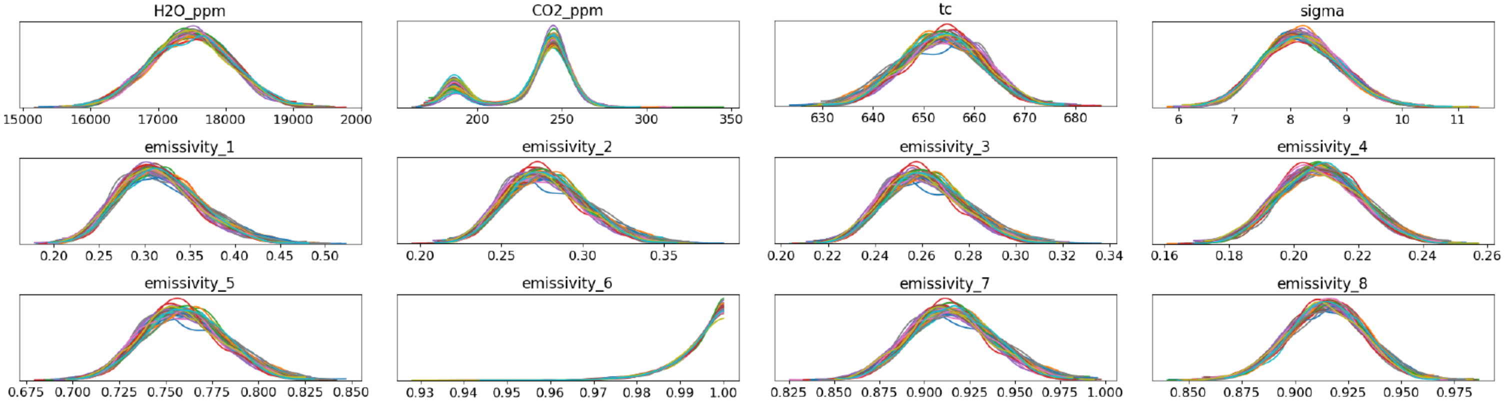


Our device: remote Temperature, Emissivity.
Range: 600-1100ºC

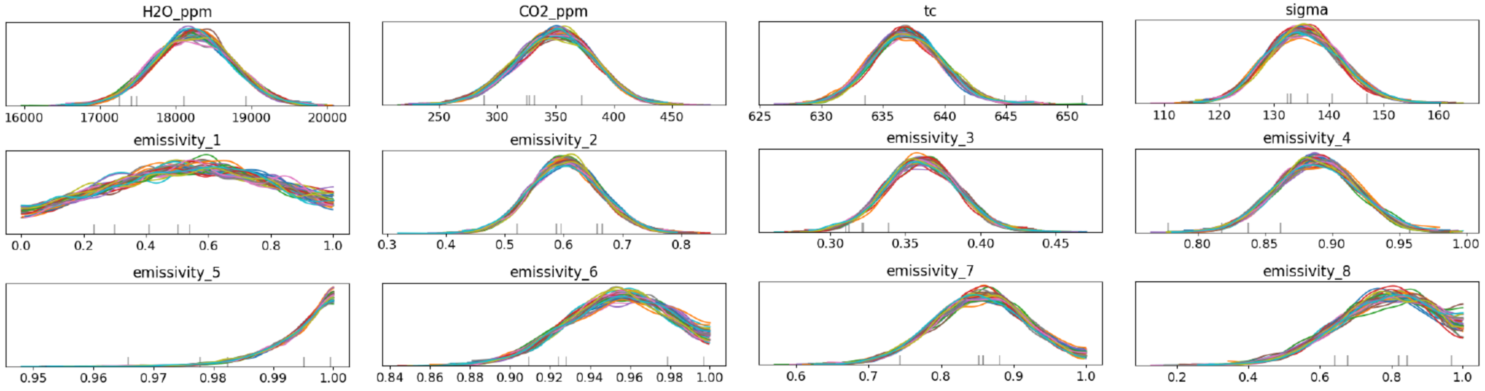
Experimental validation

Posterior probabilities

Alumina



Boron nitride

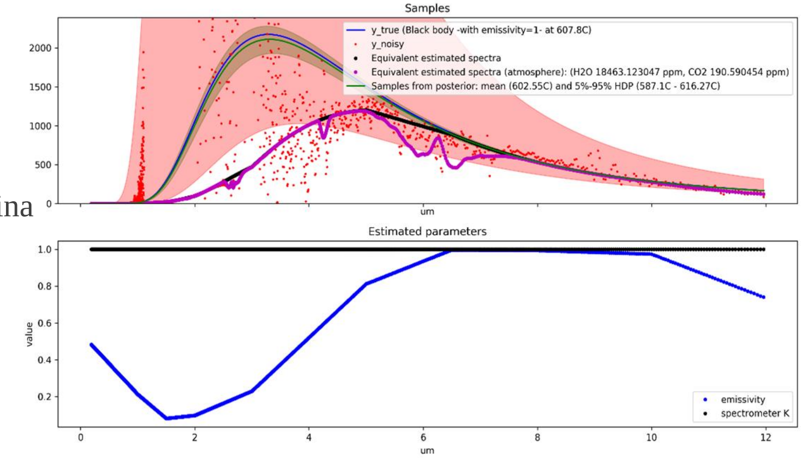


4-TES

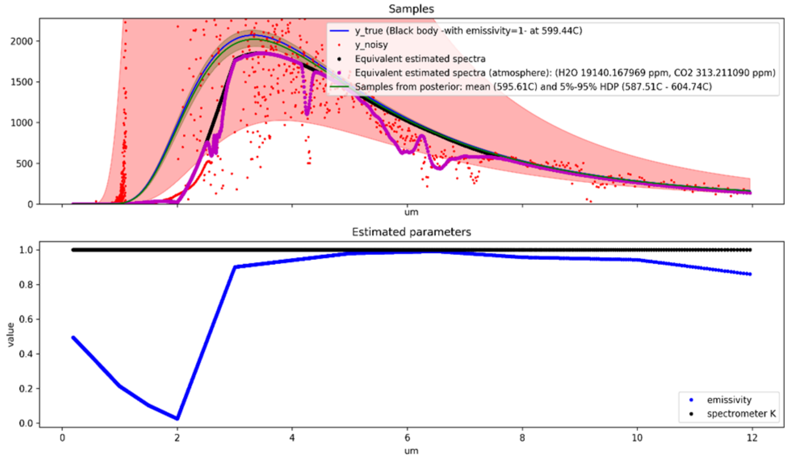
Experimental validation

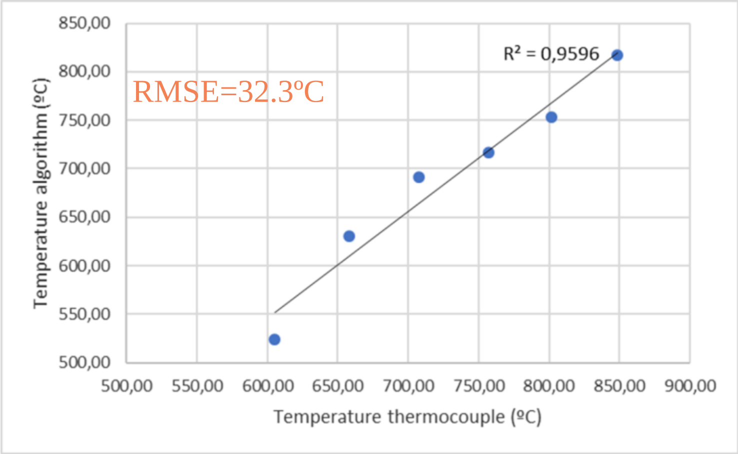
Results (radiances, spectral emissivities, temperatures)

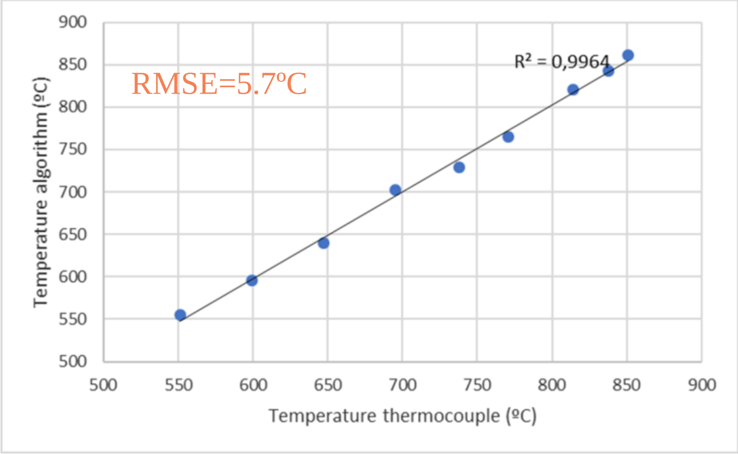
Alumina



Boron nitride





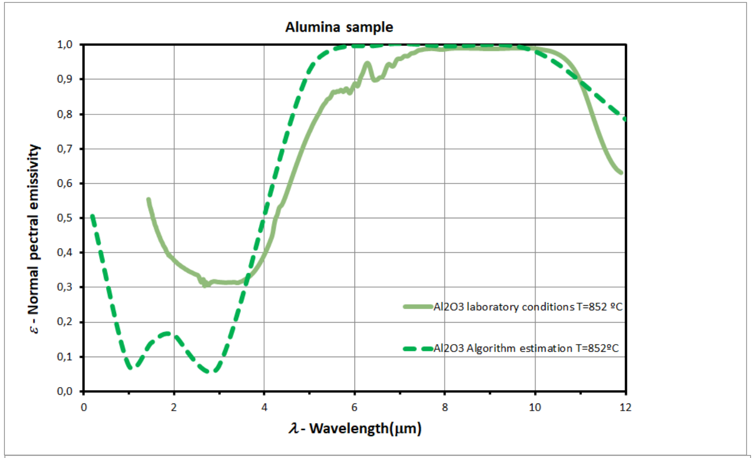


4-TES

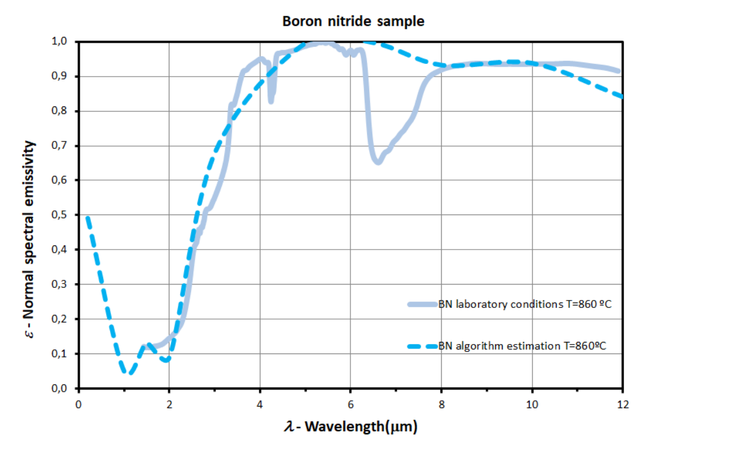
Experimental validation

Results (spectral emissivity)

Alumina



Boron nitride



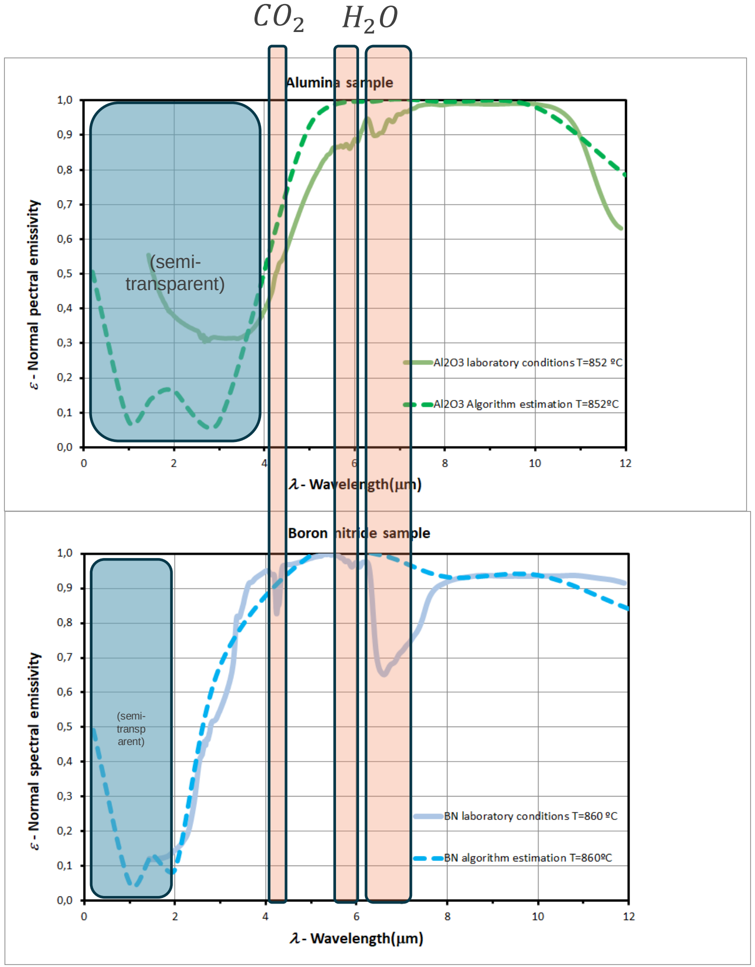
Laboratory [–] vs ours [--]

Experimental validation

Results (spectral emissivity)

Alumina

Boron nitride



Laboratory [—] vs ours [--]

Takeaways

- **Device** and **model** yielding simultaneous remote estimates for **temperature** and **spectral emissivity** of hot radiative samples in **near-steel factory conditions**.
- MCMC-based **full-probability estimates** for various process variables: $T_{bb}, (\varepsilon_1, \ldots \varepsilon_M), (x_{CO_2}, x_{H_2O})$
- **Validated** for two solid samples [600-850ºC]
- Necessary first step for remote online estimation of **slag composition** on EAF

Publication A. Picon*, **A. Alvarez-Gila***, J. A. Arteche, G. A. López, and A. Vicente, *"A Probabilistic Model and Capturing Device for Remote Simultaneous Estimation of Spectral Emissivity and Temperature of Hot Emissive Materials,"* IEEE Access, 2021.
***Equal contribution**

Patent A. Picon, **A. Alvarez-Gila,** A. Vicente, and Arteche, Jose Antonio, *"System and method for determining the emitting temperature and emissivity in a wavelength range of metallurgical products,"* PCT/IB2019/061335 , filed December 24, 2019

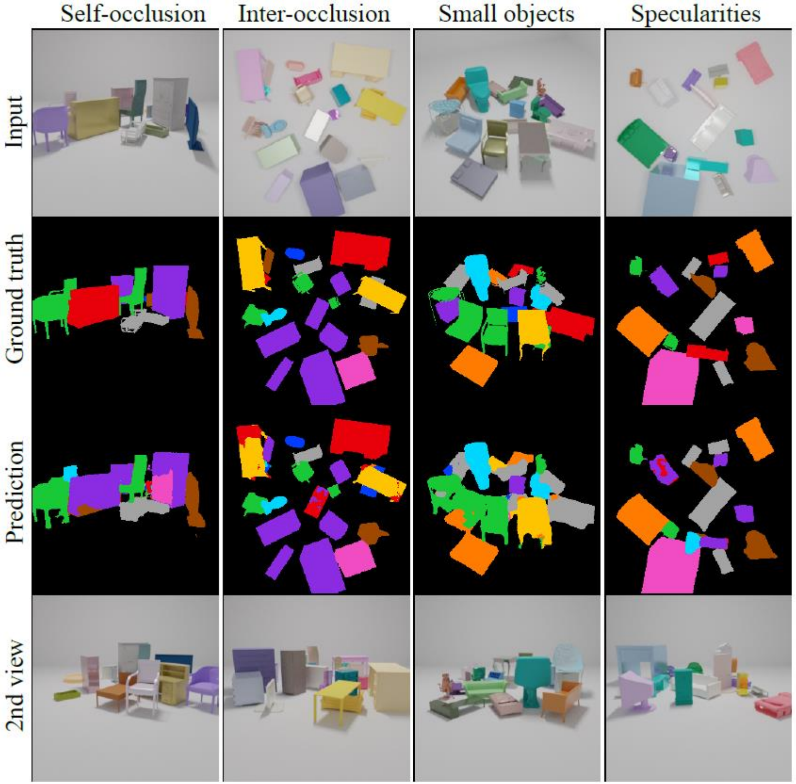
Self-supervised learning for image-to-image translation in the small data regime

Introduction

Motivation

Performance of monocular 2D semantic segmentation systems in densely populated scenes is hindered by:

- Self/inter-occlusions
- Small apparent sized objects
- Ambiguous views + fine grained categories
- Ambiguities induced by appearance variation across views (e.g. specularities)



Hypothesis:

“Data driven dense prediction models could benefit from complementary information in multi-view setups”

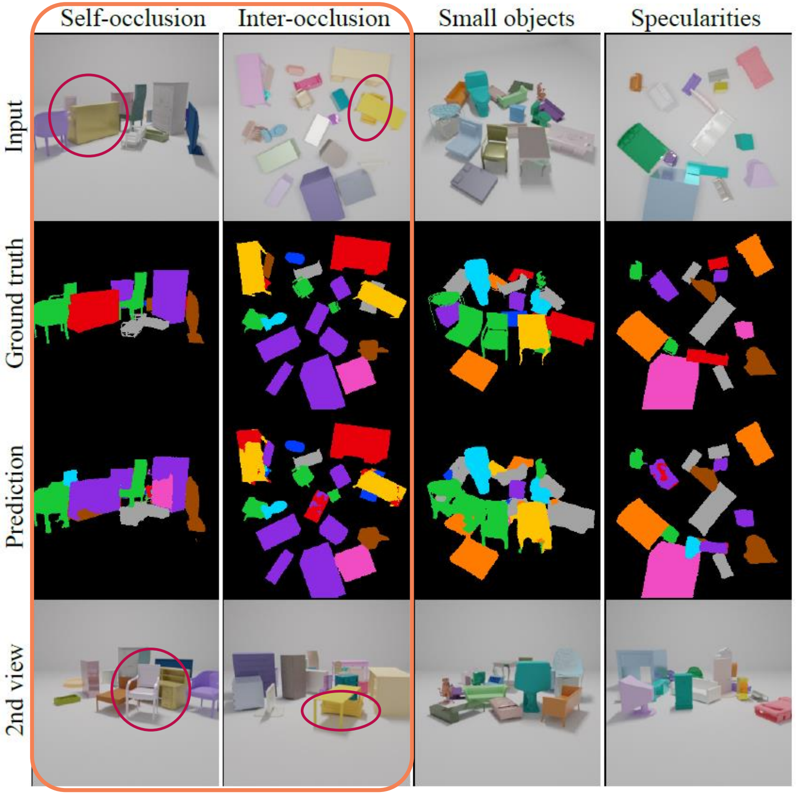
Condition: wide baselines between views (more informative)

Introduction

Motivation

Performance of monocular 2D semantic segmentation systems in densely populated scenes is hindered by:

- Self/inter-occlusions
- Small apparent sized objects
- Ambiguous views + fine grained categories
- Ambiguities induced by appearance variation across views (e.g. specularities)



Hypothesis:

“Data driven dense prediction models could benefit from complementary information in multi-view setups”

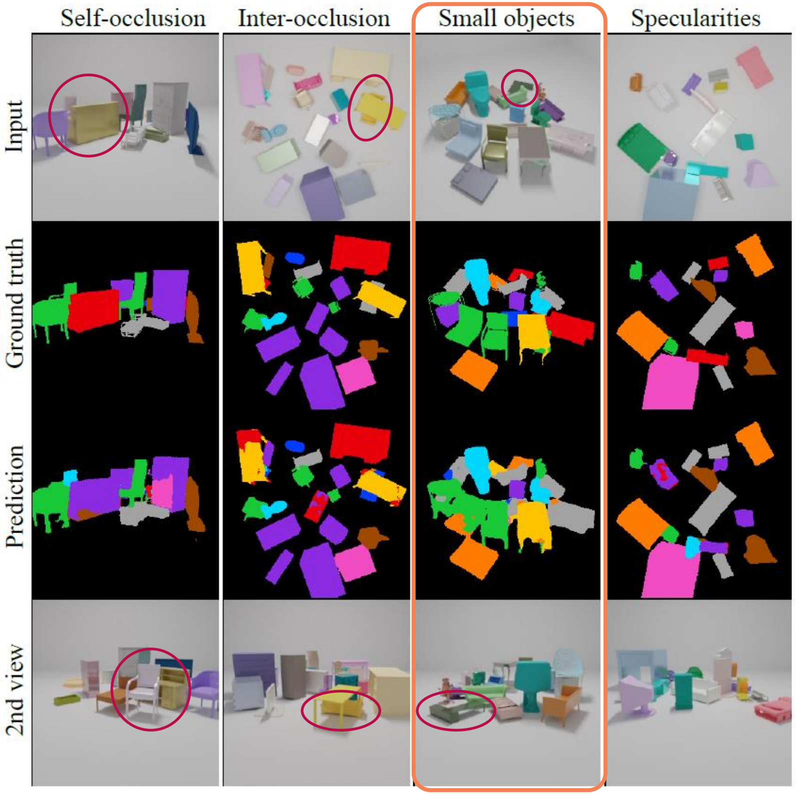
Condition: wide baselines between views (more informative)

Introduction

Motivation

Performance of monocular 2D semantic segmentation systems in densely populated scenes is hindered by:

- Self/inter-occlusions
- Small apparent sized objects
- Ambiguous views + fine grained categories
- Ambiguities induced by appearance variation across views (e.g. specularities)



Hypothesis:

“Data driven dense prediction models could benefit from complementary information in multi-view setups”

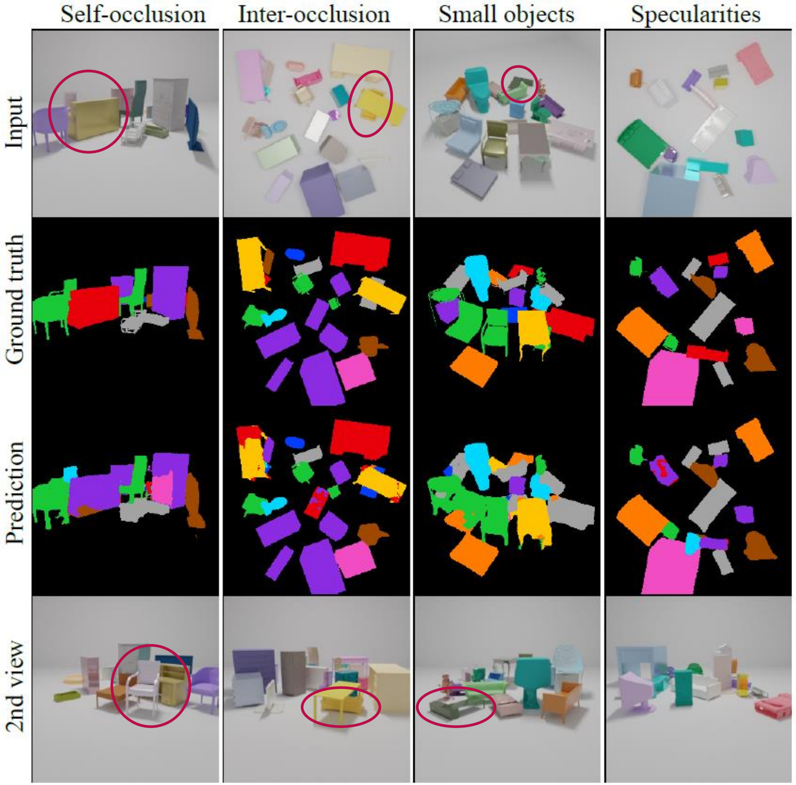
Condition: wide baselines between views (more informative)

Introduction

Motivation

Performance of monocular 2D semantic segmentation systems in densely populated scenes is hindered by:

- Self/inter-occlusions
- Small apparent sized objects
- Ambiguous views + fine grained categories
- Ambiguities induced by appearance variation across views (e.g. specularities)



Hypothesis:

"Data driven dense prediction models could benefit from complementary information in multi-view setups"

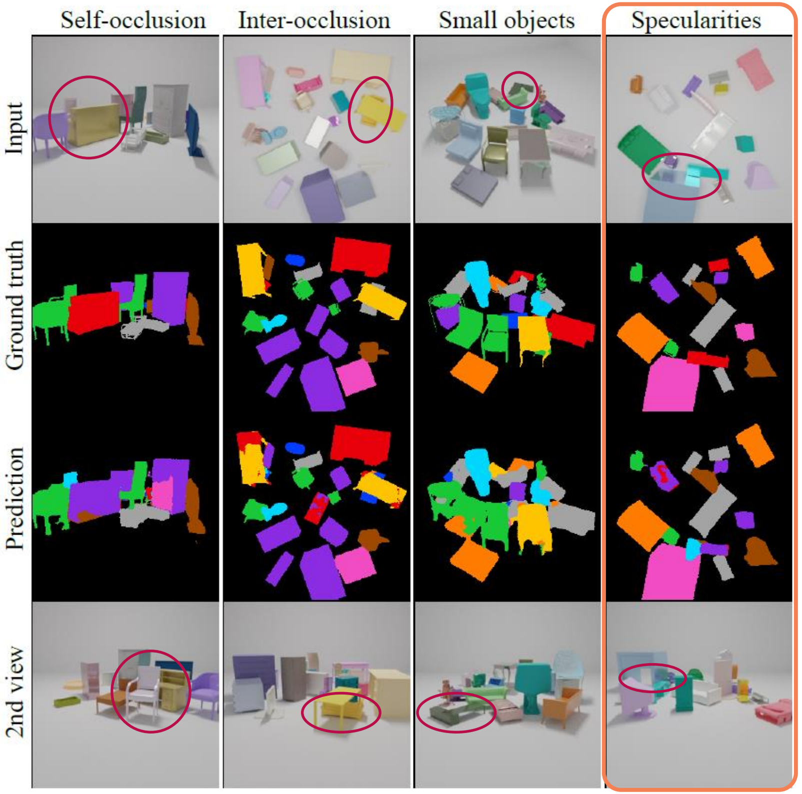
Condition: wide baselines between views (more informative)

Introduction

Motivation

Performance of monocular 2D semantic segmentation systems in densely populated scenes is hindered by:

- Self/inter-occlusions
- Small apparent sized objects
- Ambiguous views + fine grained categories
- Ambiguities induced by appearance variation across views (e.g. specularities)



Hypothesis:

"Data driven dense prediction models could benefit from complementary information in multi-view setups"

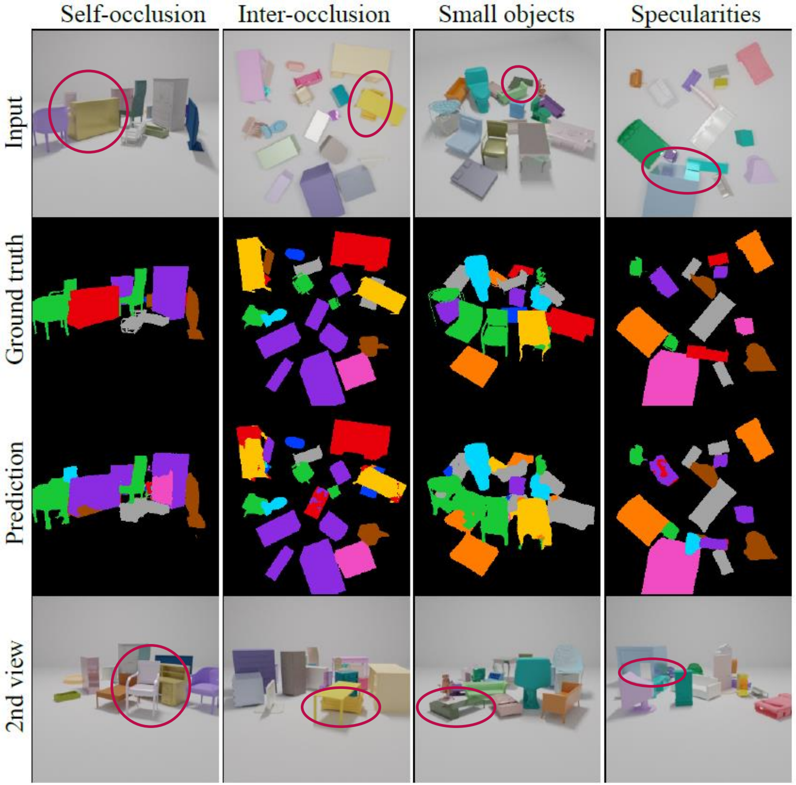
Condition: wide baselines between views (more informative)

Introduction

Motivation

Performance of monocular 2D semantic segmentation systems in densely populated scenes is hindered by:

- Self/inter-occlusions
- Small apparent sized objects
- Ambiguous views + fine grained categories
- Ambiguities induced by appearance variation across views (e.g. specularities)



Hypothesis:

"Data driven dense prediction models could benefit from complementary information in multi-view setups"

Condition: wide baselines between views (more informative)

Related work

Dataset	Wide Baseline	Object Density	Representation	Photorealism	# Scenes	# Views	# Classes
Human3.6M	Yes	Low (1)	2D images	Real	900,000 in 165 sequences	4	24
3Dpeople	Yes	Low (1)	3DM→2D	S: High B: Low	616,000 in 5,600 sequences	4	8(clothes)/14(body)
SYNTHIA	No	N/A	3DM→2D	Low	51,000 in 51 sequences	8	13
ScanNet	⋆	Low	2D→3DS	High	1.5k	⋆	40
House3D	⋆	Low	3DVE	Low	45.6k	⋆	80
Gibson	⋆	Low	3DVE	High (IBR/PCR)	1.4k	⋆	40
CARLA	⋆	⋆	3DVE	Mid-High (RT)	⋆	⋆	12

⋆ Needs to be placed/configured/generated by user; images are not readily available.

Related work

Dataset	Wide Baseline	Object Density	Representation	Photorealism	# Scenes	# Views	# Classes
Human3.6M	Yes	Low (1)	2D images	Real	900,000 in 165 sequences	4	24
3Dpeople	Yes	Low (1)	3DM→2D	S: High B: Low	616,000 in 5,600 sequences	4	8(clothes)/14(body)
SYNTHIA	No	N/A	3DM→2D	Low	51,000 in 51 sequences	8	13
ScanNet	⋆	Low	2D→3DS	High	1.5k	⋆	40
House3D	⋆	Low	3DVE	Low	45.6k	⋆	80
Gibson	⋆	Low	3DVE	High (IBR/PCR)	1.4k	⋆	40
CARLA	⋆	⋆	3DVE	Mid-High (RT)	⋆	⋆	12

⋆ Needs to be placed/configured/generated by user; images are not readily available.

Related work

Dataset	Wide Baseline	Object Density	Representation	Photorealism	# Scenes	# Views	# Classes
Human3.6M	Yes	Low (1)	2D images	Real	900,000 in 165 sequences	4	24
3Dpeople	Yes	Low (1)	3DM→2D	S: High B: Low	616,000 in 5,600 sequences	4	8(clothes)/14(body)
SYNTHIA	No	N/A	3DM→2D	Low	51,000 in 51 sequences	8	13
ScanNet	⋆	Low	2D→3DS	High	1.5k	⋆	40
House3D	⋆	Low	3DVE	Low	45.6k	⋆	80
Gibson	⋆	Low	3DVE	High (IBR/PCR)	1.4k	⋆	40
CARLA	⋆	⋆	3DVE	Mid-High (RT)	⋆	⋆	12
MVMO (ours)	Yes	High (15-20)	3DM→2D	High (PT, UOM)	116k (uncorrelated)	25	11

⋆ Needs to be placed/configured/generated by user; images are not readily available.

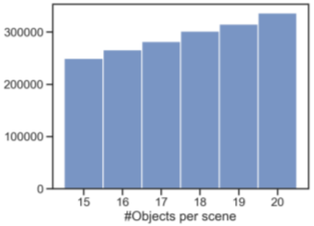
MVMO

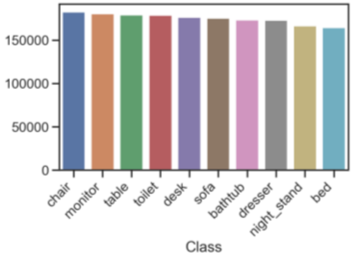
The Multi-View, Multi-Object Dataset

100k train + 8k val + 8k test scenes

15-20 objects/scene:

- Sampled from ModelNet10
- 10 annotated categories
- Diversified appearance/BSDF

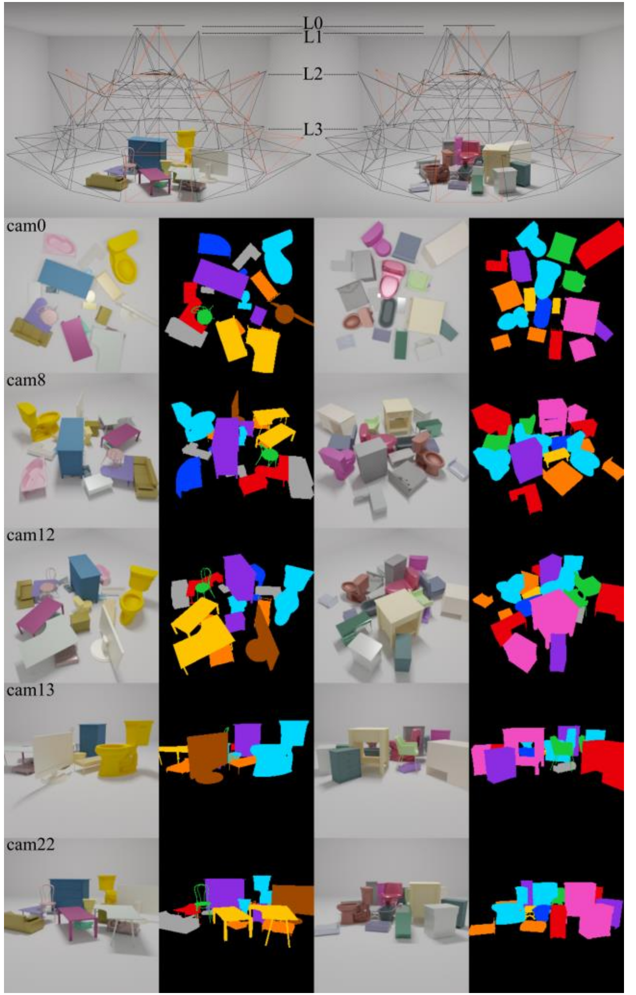




25 camera locations at the upper hemisphere at 4 levels

Wide-baselines →large disparities

High density of objects →Multiple occlusions



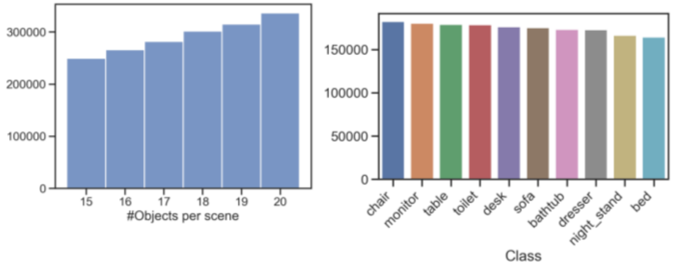
MVMO

The Multi-View, Multi-Object Dataset

100k train + 8k val + 8k test scenes

15-20 objects/scene:

- Sampled from ModelNet10
- 10 annotated categories
- Diversified appearance/BSDF



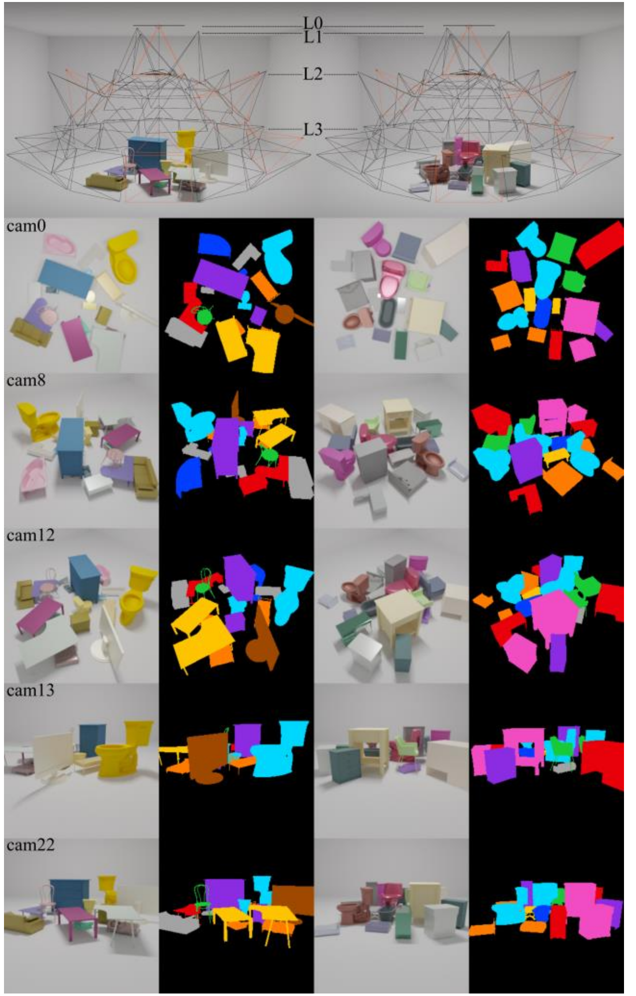
25 camera locations at the upper hemisphere at 4 levels

Wide-baselines →large disparities

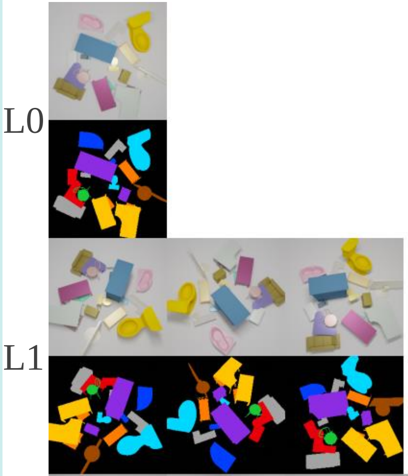
High density of objects →Multiple occlusions

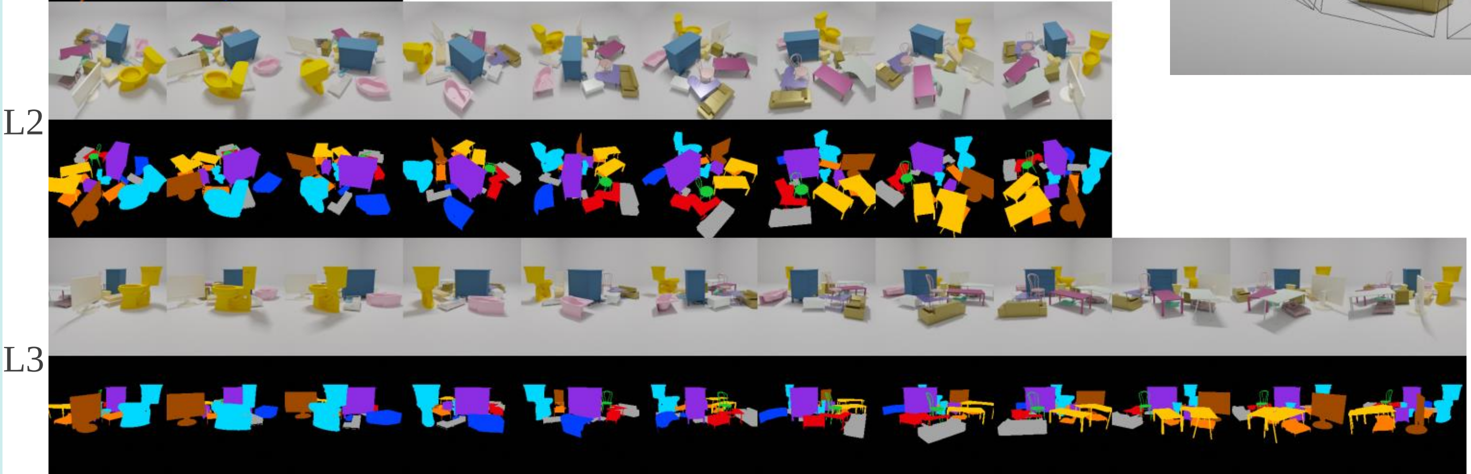
Goal: boost research in [wide baseline]…

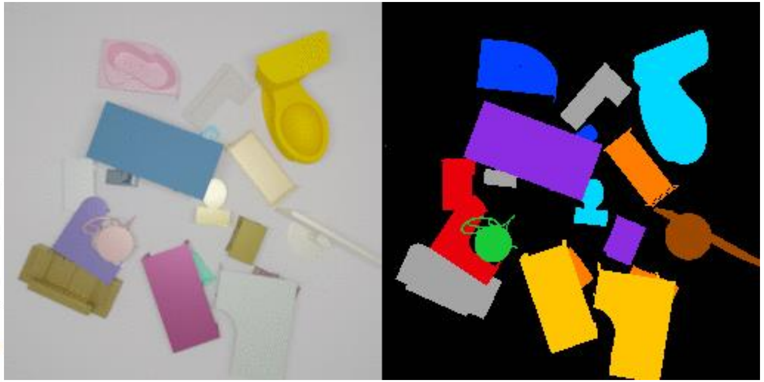
- (i) **multi-view semantic segmentation**
- (ii) **cross-view semantic transfer** from single view labels.



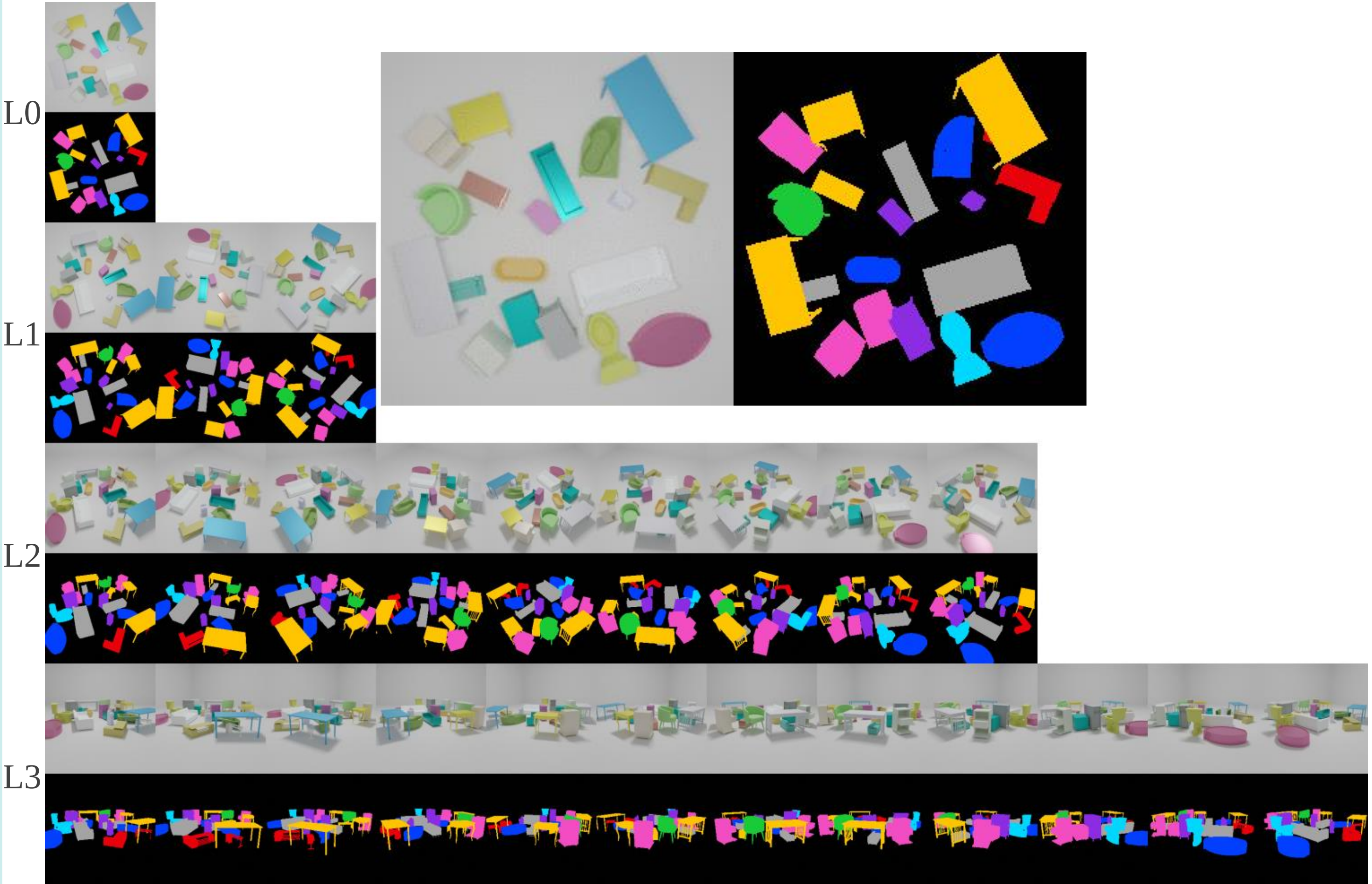
Scene samples



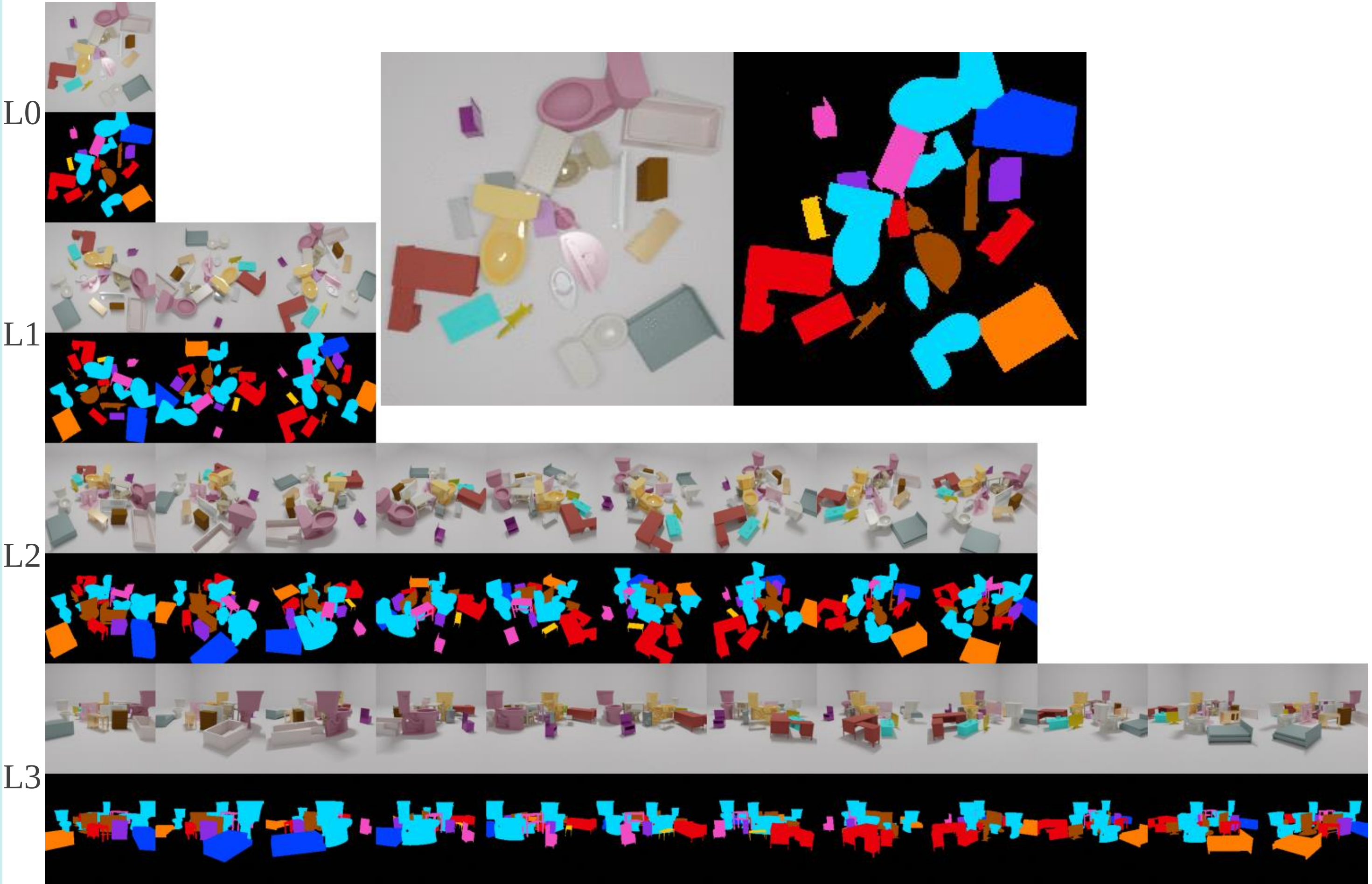




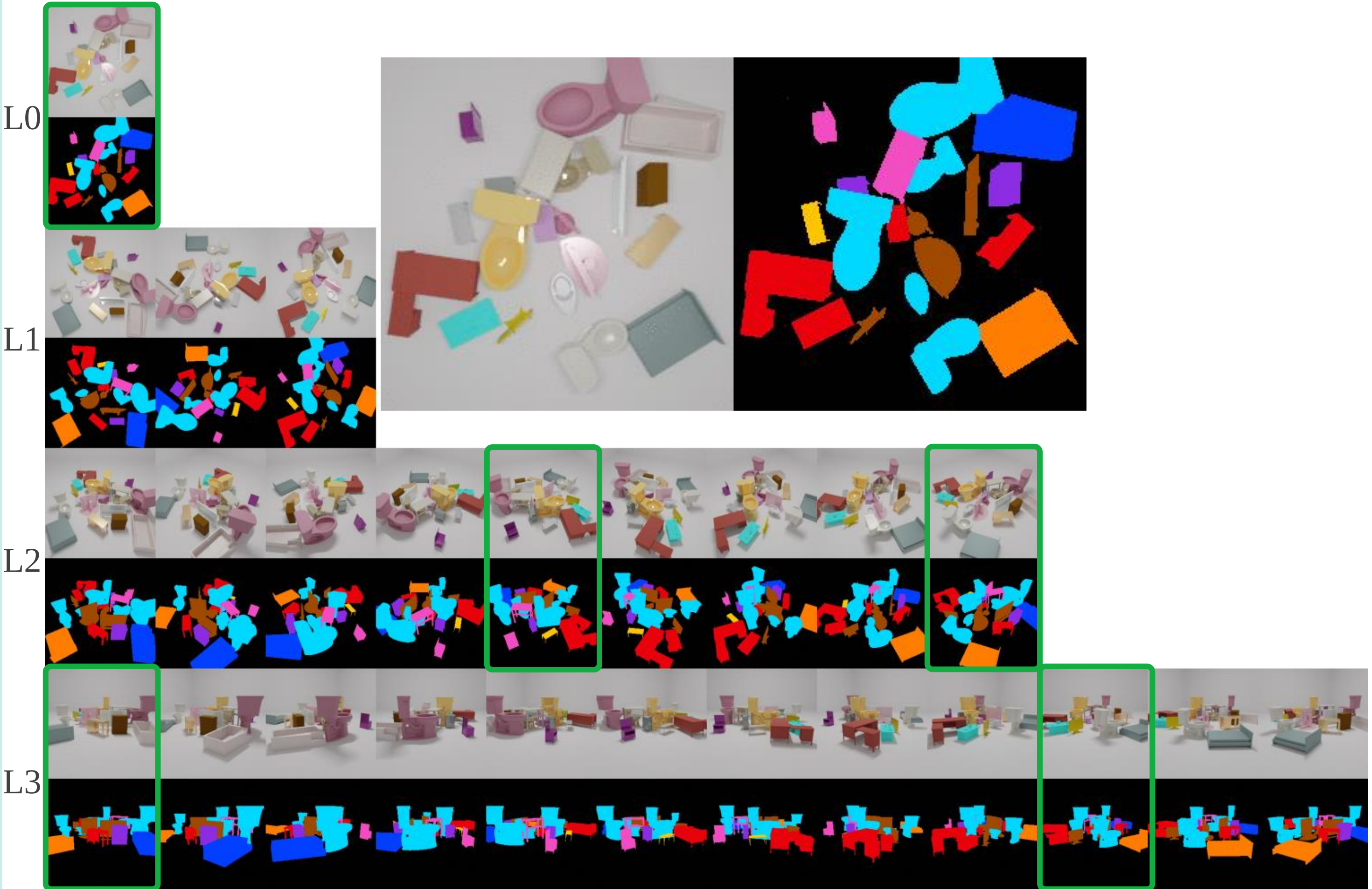
Scene samples



Scene samples



Scene samples

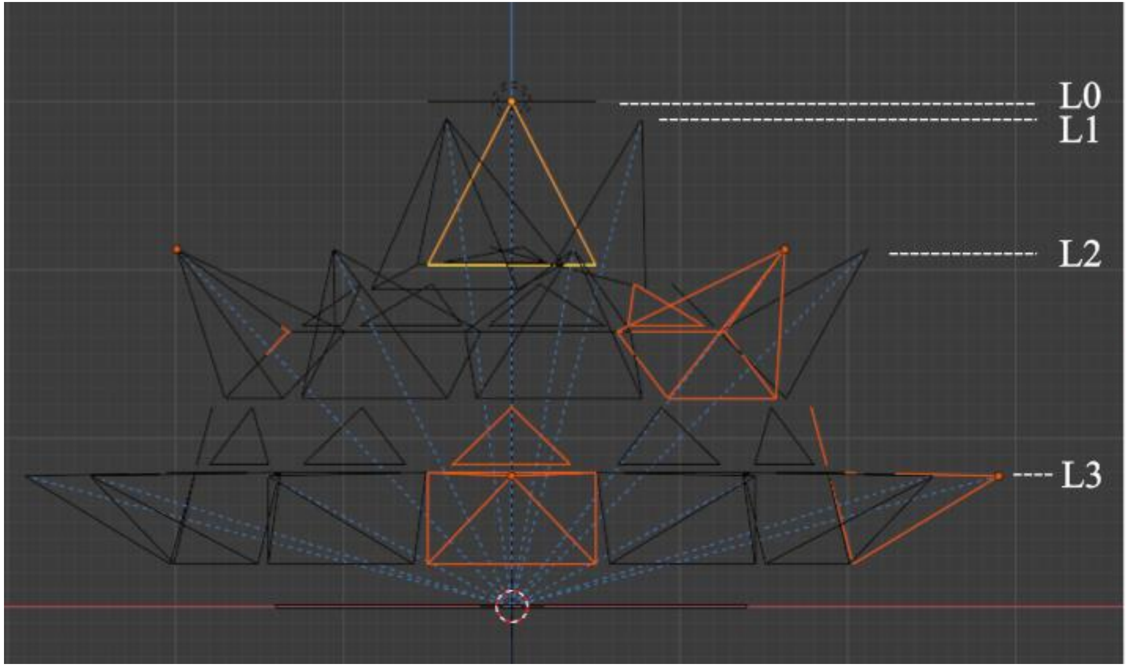


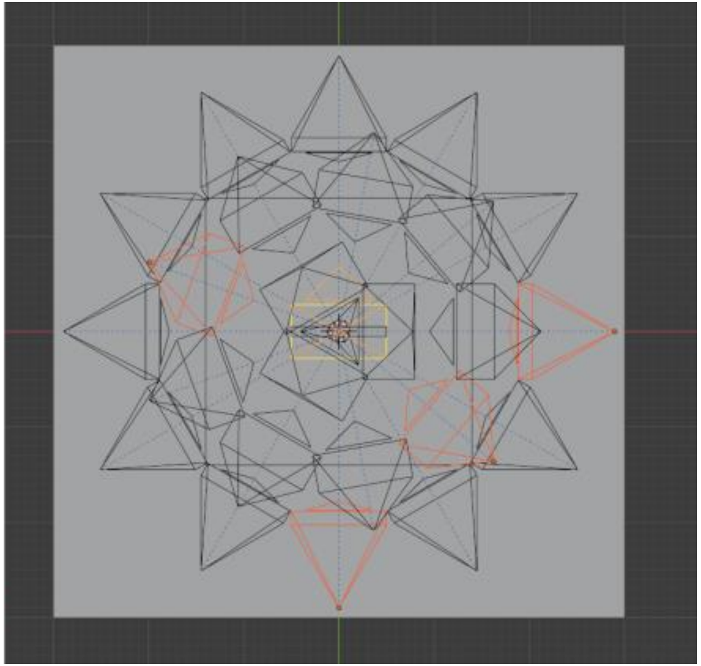
Experimental baselines

Setup

5 cameras at 3 different levels (L0, L2, L3)

Imagenet pretrained U-Net architecture





Experimental baselines

Experiment 1: cross view semantic transfer via direct testing

Given a model, $f_{v_r \rightarrow ss_r}$, trained on (v_r, ss_r) pairs, we want to feed it with inputs from view v_t and obtain ss_t segmentation results referenced to v_t (ss_t).

Subset	test (v_t) \ train (v_r)	cam0	cam8	cam12	cam13	cam22
Other objs.	L0.cam0	**71.12**	29.09	29.61	14.28	14.88
	L2.cam8	24.63	**70.21**	**70.16**	28.14	28.54
	L2.cam12	25.14	69.09	70.05	27.73	28.29
	L3.cam13	12.18	31.26	31.46	**59.18**	58.72
	L3.cam22	12.11	30.10	30.59	58.39	**59.41**
Same objs.	L0.cam0	**80.55**	29.92	29.69	14.00	14.51
	L2.cam8	27.11	**77.90**	77.71	27.24	27.46
	L2.cam12	28.01	76.87	**77.97**	26.94	27.52
	L3.cam13	12.90	32.16	32.29	**65.87**	65.69
	L3.cam22	12.76	31.00	31.68	64.84	**66.09**

Experimental baselines

Experiment 1: cross view semantic transfer via direct testing

Given a model, $f_{v_r \rightarrow ss_r}$, trained on (v_r, ss_r) pairs, we want to feed it with inputs from view v_t and obtain ss_t segmentation results referenced to v_t (ss_t).

Subset	test (v_t) \ train (v_r)	cam0	cam8	cam12	cam13	cam22
Other objs.	L0.cam0	**71.12**	29.09	29.61	14.28	14.88
	L2.cam8	24.63	**70.21**	**70.16**	28.14	28.54
	L2.cam12	25.14	69.09	70.05	27.73	28.29
	L3.cam13	12.18	31.26	31.46	**59.18**	58.72
	L3.cam22	12.11	30.10	30.59	58.39	**59.41**
Same objs.	L0.cam0	**80.55**	29.92	29.69	14.00	14.51
	L2.cam8	27.11	**77.90**	77.71	27.24	27.46
	L2.cam12	28.01	76.87	**77.97**	26.94	27.52
	L3.cam13	12.90	32.16	32.29	**65.87**	65.69
	L3.cam22	12.76	31.00	31.68	64.84	**66.09**

Fully-sup monocular

Experimental baselines

Experiment 1: cross view semantic transfer via direct testing

Given a model, $f_{v_r \rightarrow ss_r}$, trained on (v_r, ss_r) pairs, we want to feed it with inputs from view v_t and obtain ss_t segmentation results referenced to v_t (ss_t).

Subset	test (v_t) \ train (v_r)	cam0	cam8	cam12	cam13	cam22
Other objs.	L0.cam0	**71.12**	29.09	29.61	14.28	14.88
	L2.cam8	24.63	**70.21**	**70.16**	28.14	28.54
	L2.cam12	25.14	69.09	70.05	27.73	28.29
	L3.cam13	12.18	31.26	31.46	**59.18**	58.72
	L3.cam22	12.11	30.10	30.59	58.39	**59.41**
Same objs.	L0.cam0	**80.55**	29.92	29.69	14.00	14.51
	L2.cam8	27.11	**77.90**	77.71	27.24	27.46
	L2.cam12	28.01	76.87	**77.97**	26.94	27.52
	L3.cam13	12.90	32.16	32.29	**65.87**	65.69
	L3.cam22	12.76	31.00	31.68	64.84	**66.09**

Good generalization to unseen objects

Experimental baselines

Experiment 1: cross view semantic transfer via direct testing

Given a model, $f_{v_r \rightarrow ss_r}$, trained on (v_r, ss_r) pairs, we want to feed it with inputs from view v_t and obtain ss_t segmentation results referenced to v_t (ss_t).

Subset	test (v_t) \ train (v_r)	cam0	cam8	cam12	cam13	cam22
Other objs.	L0.cam0	**71.12**	29.09	29.61	14.28	14.88
	L2.cam8	24.63	**70.21**	**70.16**	28.14	28.54
	L2.cam12	25.14	69.09	70.05	27.73	28.29
	L3.cam13	12.18	31.26	31.46	**59.18**	58.72
	L3.cam22	12.11	30.10	30.59	58.39	**59.41**
Same objs.	L0.cam0	**80.55**	29.92	29.69	14.00	14.51
	L2.cam8	27.11	**77.90**	77.71	27.24	27.46
	L2.cam12	28.01	76.87	**77.97**	26.94	27.52
	L3.cam13	12.90	32.16	32.29	**65.87**	65.69
	L3.cam22	12.76	31.00	31.68	64.84	**66.09**

Good generalization to unseen objects

L2↔L2
No gap

Experimental baselines

Experiment 1: cross view semantic transfer via direct testing

Given a model, $f_{v_r \to ss_r}$, trained on (v_r, ss_r) pairs, we want to feed it with inputs from view v_t and obtain ss_t segmentation results referenced to v_t (ss_t).

Subset	test (v_t) \ train (v_r)	cam0	cam8	cam12	cam13	cam22
Other objs.	L0.cam0	**71.12**	29.09	29.61	14.28	14.88
	L2.cam8	24.63	**70.21**	**70.16**	28.14	28.54
	L2.cam12	25.14	69.09	70.05	27.73	28.29
	L3.cam13	12.18	31.26	31.46	**59.18**	58.72
	L3.cam22	12.11	30.10	30.59	58.39	**59.41**
Same objs.	L0.cam0	**80.55**	29.92	29.69	14.00	14.51
	L2.cam8	27.11	**77.90**	77.71	27.24	27.46
	L2.cam12	28.01	76.87	**77.97**	26.94	27.52
	L3.cam13	12.90	32.16	32.29	**65.87**	65.69
	L3.cam22	12.76	31.00	31.68	64.84	**66.09**

Good generalization to unseen objects

L2↔L2
No gap

L0↔L2
Performance drop

Experimental baselines

Experiment 1: cross view semantic transfer via direct testing

Given a model, $f_{v_r \to ss_r}$, trained on (v_r, ss_r) pairs, we want to feed it with inputs from view v_t and obtain ss_t segmentation results referenced to v_t (ss_t).

Subset	test (v_t) \ train (v_r)	cam0	cam8	cam12	cam13	cam22
Other objs.	L0.cam0	**71.12**	29.09	29.61	14.28	14.88
	L2.cam8	24.63	**70.21**	**70.16**	28.14	28.54
	L2.cam12	25.14	69.09	70.05	27.73	28.29
	L3.cam13	12.18	31.26	31.46	**59.18**	58.72
	L3.cam22	12.11	30.10	30.59	58.39	**59.41**
Same objs.	L0.cam0	**80.55**	29.92	29.69	14.00	14.51
	L2.cam8	27.11	**77.90**	77.71	27.24	27.46
	L2.cam12	28.01	76.87	**77.97**	26.94	27.52
	L3.cam13	12.90	32.16	32.29	**65.87**	65.69
	L3.cam22	12.76	31.00	31.68	64.84	**66.09**

Good generalization to unseen objects

L2↔L2
No gap

L0↔L2
Performance drop

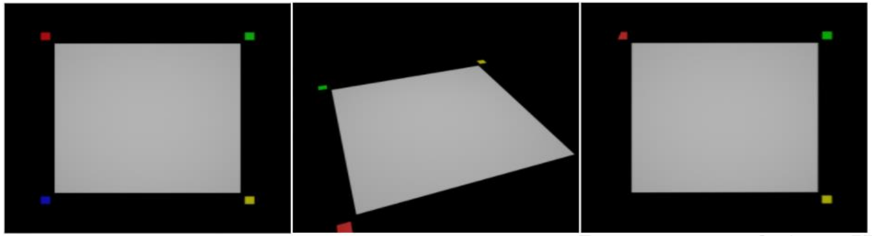
Experimental baselines

Experiment 2: planar homography-based transfer

Given a model, $f_{v_r \rightarrow ss_r}$, trained on (v_r, ss_r) pairs, we want to feed it with inputs from view v_t and obtain ss_t segmentation results referenced to v_t (ss_t).

Planar (3x3) homography mapping holds only for (i) quasi-planar scenes (ii) distant objects

Computation (4-point correspondence):



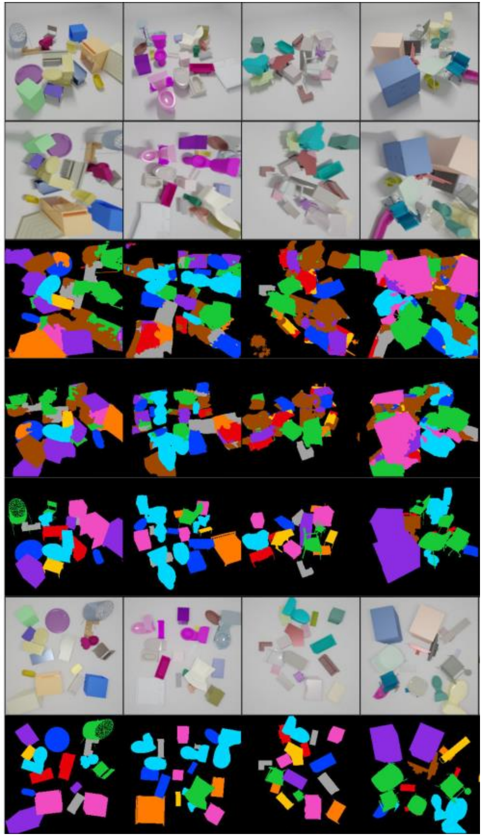
v_r =L0.cam0 v_t =L2.cam8 Reconstructed $\hat{v_r} = H^{z=0}_{t \rightarrow r}(v_t)$

	v_r =L0.cam0→ v_t =L2.cam8		v_r =L2.cam8→ v_t =L0.cam0	
	Other objs.	Same objs.	Other objs.	Same objs.
IoU	28.72 (+4,09)	31.29	24.35 (-4,74)	24.84

Experimental baseline

Experiment 2: planar homography-based transfer

v_r =L0.cam0$\rightarrow$ v_t =L2.cam8



a) v_t, input at inference time.

b) Planar homography-based estimate of v_r: $\hat{v_r} = H^{z=0}_{t\rightarrow r}(v_t)$

c) Predicted ss_r: $s\hat{s}_r = f_{v_r \rightarrow ss_r}(\hat{v_r})$

d) Predicted ss_t: $s\hat{s}_t = H^{z=0}_{r\rightarrow t}(s\hat{s}_r)$, where $H^{z=0}_{r\rightarrow t} = (H^{z=0}_{t\rightarrow r})^{-1}$.

e) Ground truth for the task, ss_t.

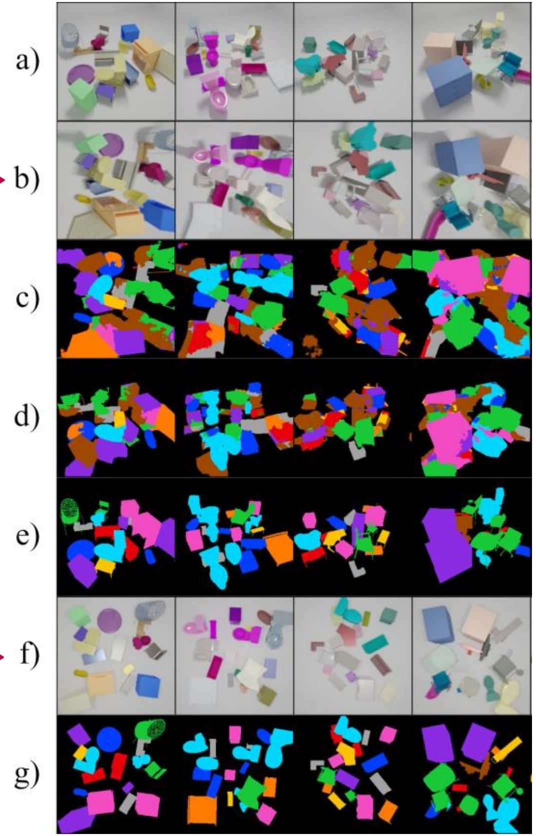
f) Ground truth v_r view. Used for training the $f_{v_r \rightarrow ss_r}$ model.

g) Ground truth ss_r semantic map. Used for training the $f_{v_r \rightarrow ss_r}$ model.

Experimental baseline

Experiment 2: planar homography-based transfer

$v_r = \text{L0.cam0} \rightarrow v_t = \text{L2.cam8}$



a) v_t, input at inference time.

b) Planar homography-based estimate of v_r: $\hat{v_r} = H^{z=0}_{t\rightarrow r}(v_t)$

c) Predicted ss_r: $s\hat{s}_r = f_{v_r \rightarrow ss_r}(\hat{v_r})$

d) Predicted ss_t: $s\hat{s}_t = H^{z=0}_{r\rightarrow t}(s\hat{s}_r)$, where $H^{z=0}_{r\rightarrow t} = (H^{z=0}_{t\rightarrow r})^{-1}$.

e) Ground truth for the task, ss_t.

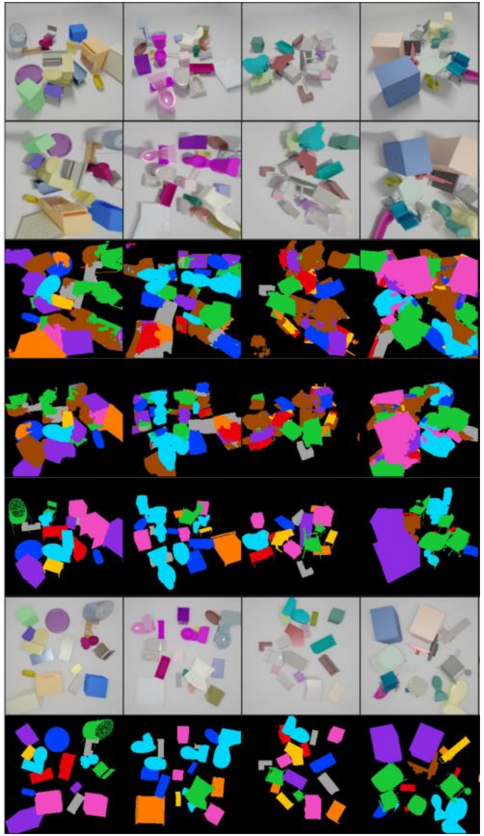
f) Ground truth v_r view. Used for training the $f_{v_r \rightarrow ss_r}$ model.

g) Ground truth ss_r semantic map. Used for training the $f_{v_r \rightarrow ss_r}$ model.

Experimental baseline

Experiment 2: planar homography-based transfer

v_r =L0.cam0$\rightarrow$ v_t =L2.cam8



a) v_t, input at inference time.

b) Planar homography-based estimate of v_r: $\hat{v_r} = H^{z=0}_{t\rightarrow r}(v_t)$

c) Predicted ss_r: $s\hat{s}_r = f_{v_r \rightarrow ss_r}(\hat{v_r})$

d) Predicted ss_t: $s\hat{s}_t = H^{z=0}_{r\rightarrow t}(s\hat{s}_r)$, where $H^{z=0}_{r\rightarrow t} = (H^{z=0}_{t\rightarrow r})^{-1}$.

e) Ground truth for the task, ss_t.

f) Ground truth v_r view. Used for training the $f_{v_r \rightarrow ss_r}$ model.

g) Ground truth ss_r semantic map. Used for training the $f_{v_r \rightarrow ss_r}$ model.

Takeaways

Introduced **MVMO**: (i) wide baseline (ii) multi-view (iii) synthetic dataset (iv) with semantic segmentation annotations that features (v) high object density and (vi) large number of occlusions.

Goal: Propel research in

(i) **multi-view semantic segmentation**

(ii) **cross-view semantic transfer**,

addressing limitations of monocular setups in heavily-occluded scenes

Publication **A. Alvarez-Gila**, J. Van De Weijer, Y. Wang, and E. Garrote, *"MVMO: A Multi-Object Dataset for Wide Baseline Multi-View Semantic Segmentation,"* ICIP 2022 (accepted).

https://aitorshuffle.github.io/projects/mvmo/

Self-supervised learning for image-to-image translation in the small data regime

Problem statement

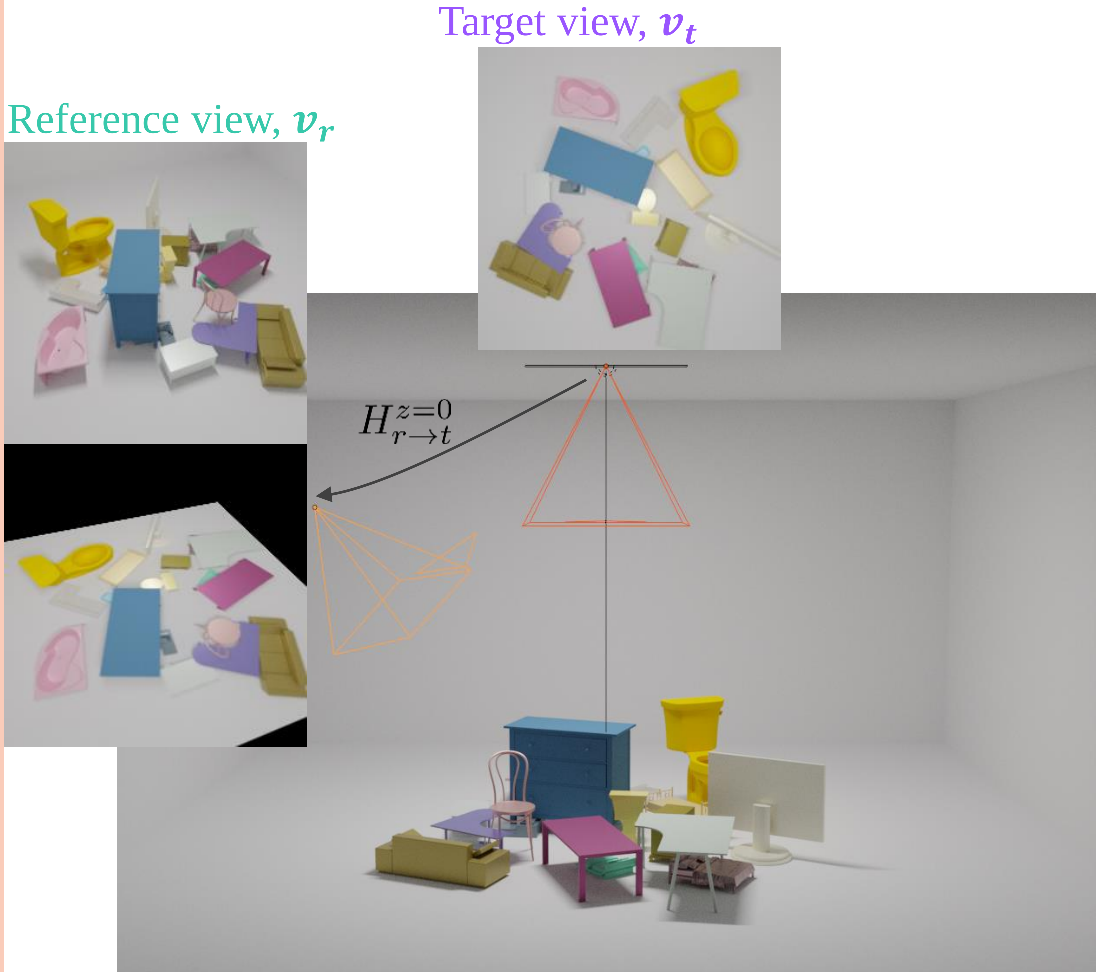
Reference view, v_{r}

- Monocular semantic segmentation system trained fully-supervised requiring camera relocation (e.g. industrial in-line production system): $v_{\mathrm{r}} \rightarrow v_t$
- Inference with model trained on reference view (v_{r}) fails (domain shift).
- Labeling from new pose is very costly.

Problem statement

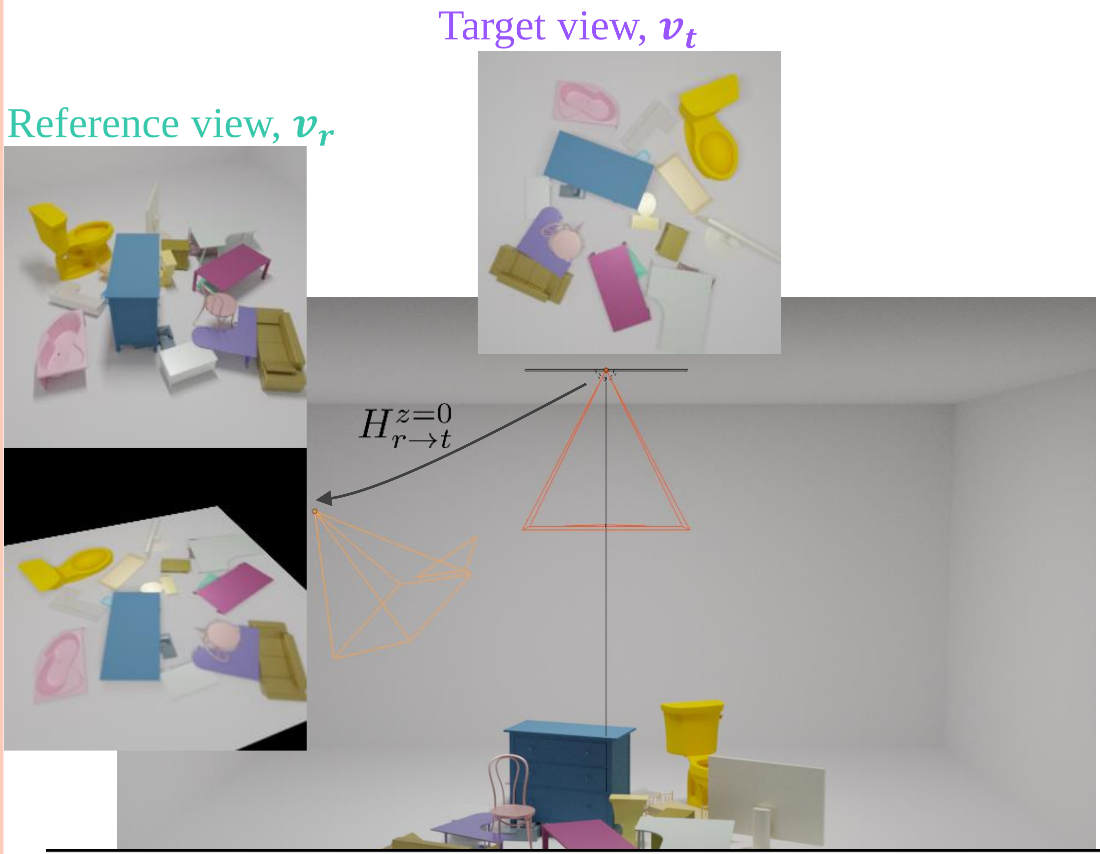
- Monocular semantic segmentation system trained fully-supervised requiring camera relocation (e.g. industrial in-line production system): $v_r \rightarrow v_t$
- Inference with model trained on reference view (v_r) fails (domain shift).
- Labeling from new pose is very costly.

Problem statement



- Monocular semantic segmentation system trained fully-supervised requiring camera relocation (e.g. industrial in-line production system): $v_r \rightarrow v_t$
- Inference with model trained on reference view (v_r) fails (domain shift).
- Labeling from new pose is very costly.
- Pure geometry-based tools (3x3 planar homography $H_{r\to t}^{z=0}$) fail for wide baselines.
- We need cross-view knowledge transfer tools that exploit statistical priors for cheap camera relocations.

Problem statement



- Monocular semantic segmentation system trained fully-supervised requiring camera relocation (e.g. industrial in-line production system): $v_r \to v_t$
- Inference with model trained on reference view (v_r) fails (domain shift).
- Labeling from new pose is very costly.
- Pure geometry-based tools (3x3 planar homography $H^{z=0}_{r \to t}$) fail for wide baselines.
- We need cross-view knowledge transfer tools that exploit statistical priors for cheap camera relocations.

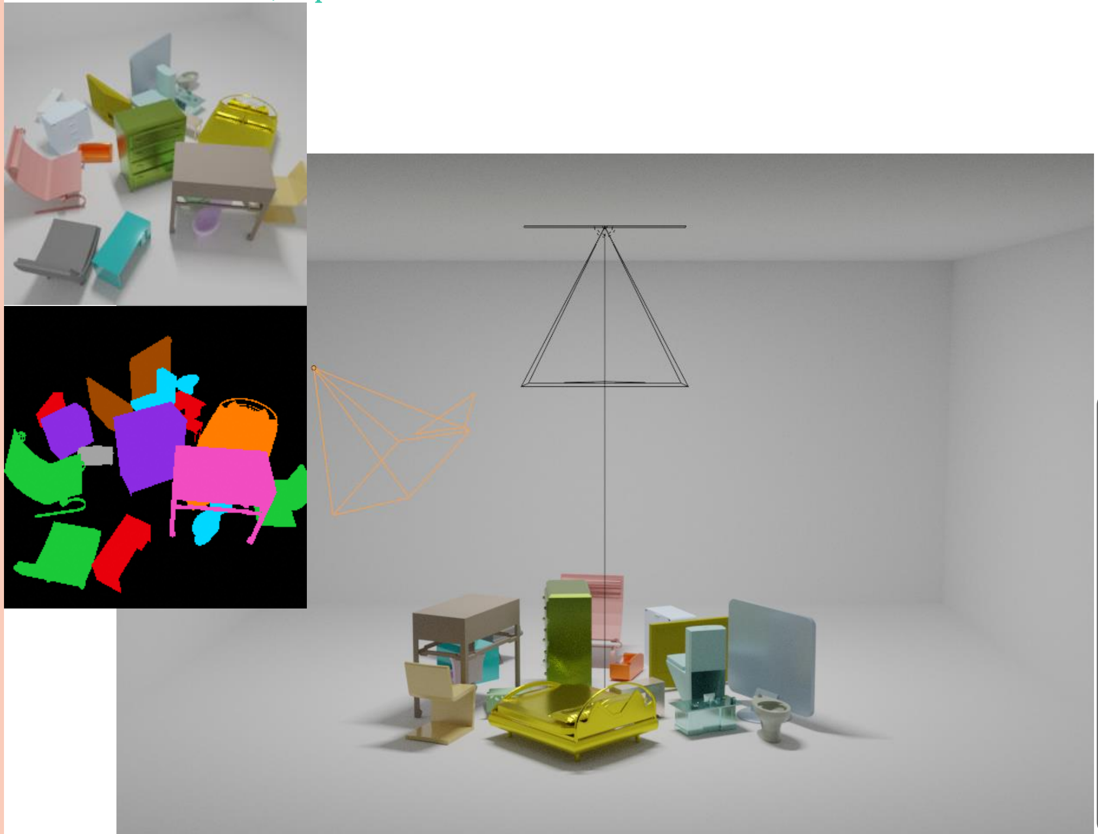
Idea:

Leverage unlabeled (reference, target) view pairs.

Chapter	Small data	Technique	Prior knowledge/Physics	Synthetic
Ch.5 (MVMO), Ch.6 (ZPCVNet)	Few labels	CNN	Path tracing	(Input, Ground Truth)

Zero pair, cross-view semantic segmentation

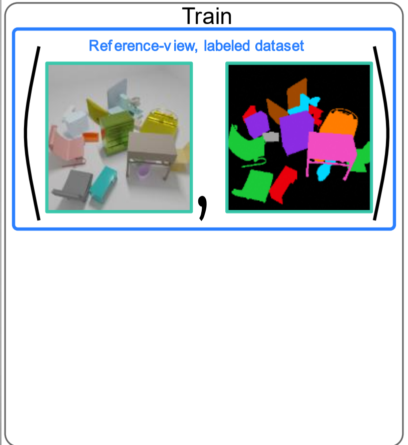
Reference view, v_r

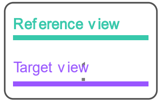


- New semi-supervised task:

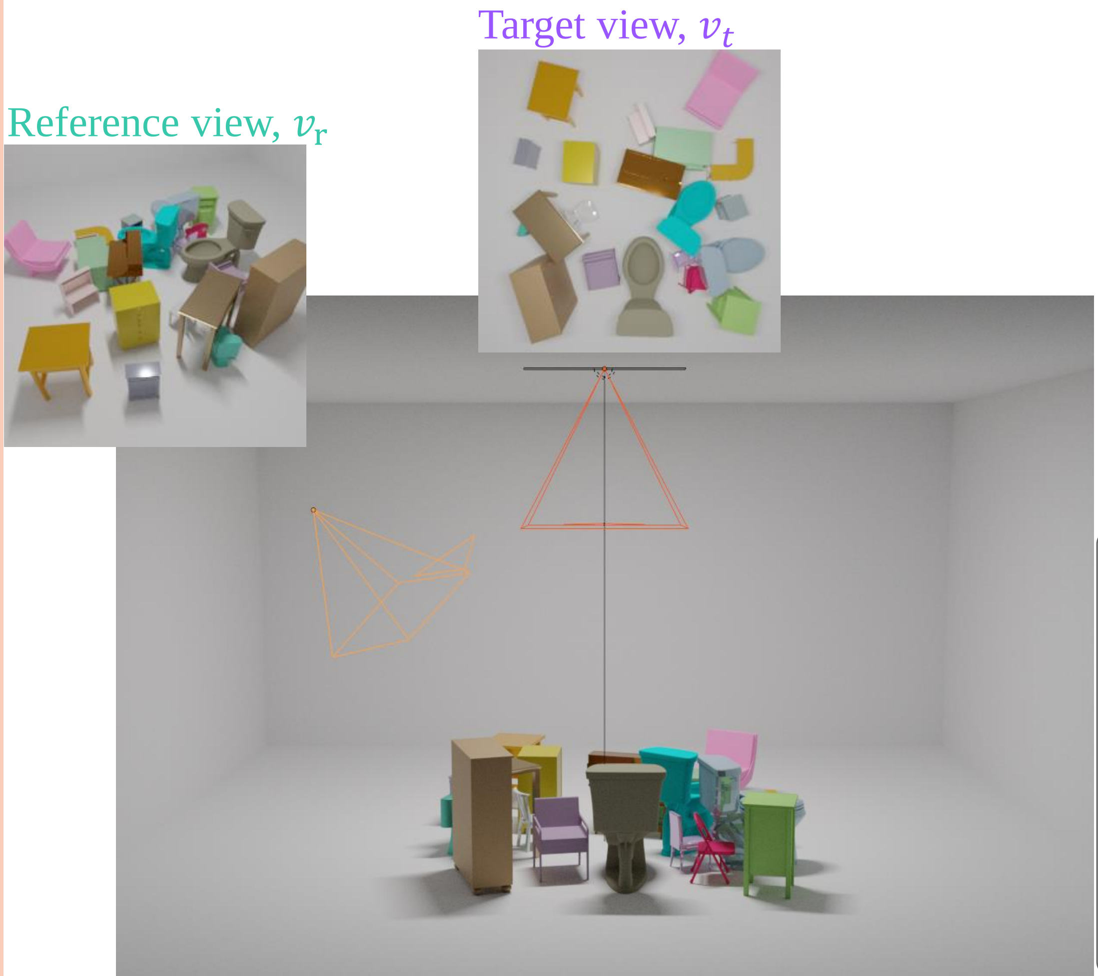
  Zero-Pair, Cross-View semantic segmentation

- Train on:
  - Reference view (v_r), labeled dataset $\mathcal{D}_{r,l}$
  - Disjoint cross-view, unlabeled dataset $\mathcal{D}_u$





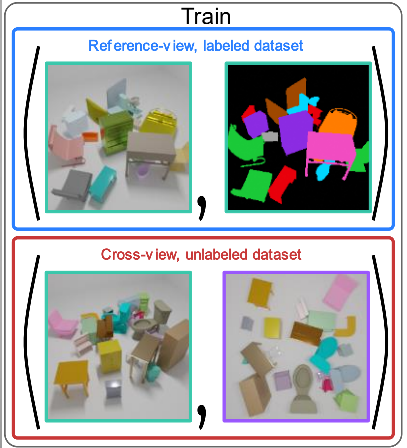
Zero pair, cross-view semantic segmentation

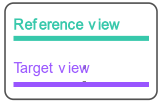


- New semi-supervised task:

  Zero-Pair, Cross-View semantic segmentation

- Train on:
  - Reference view (v_{r}), labeled dataset $\mathcal{D}_{r,l}$
  - Disjoint cross-view, unlabeled dataset $\mathcal{D}_u$





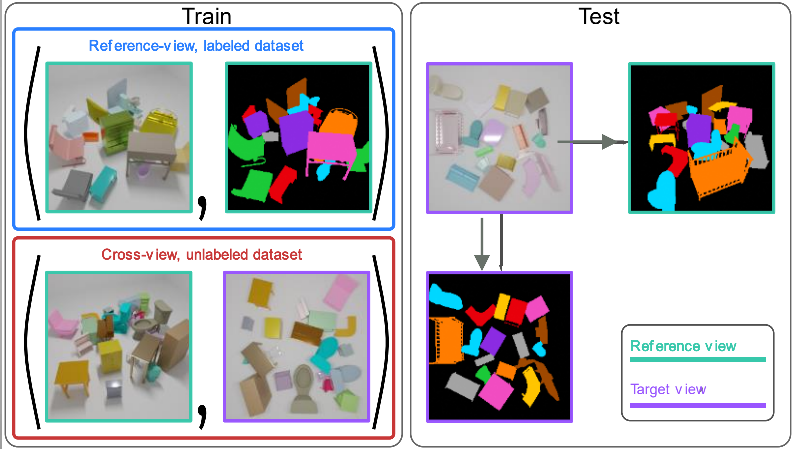
Zero pair, cross-view semantic segmentation

Target view, v_t

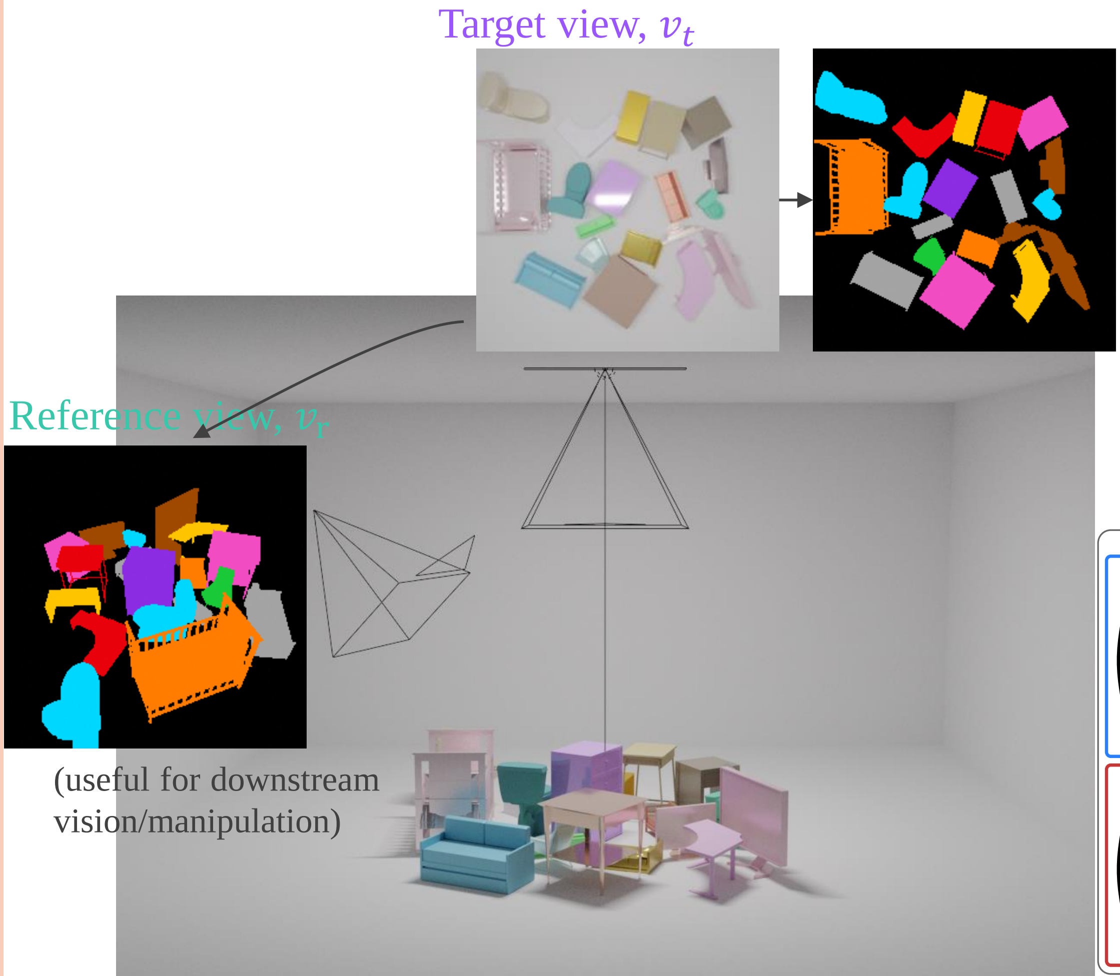
- New semi-supervised task:

  Zero-Pair, Cross-View semantic segmentation

- Train on:
  - Reference view (v_r), labeled dataset $\mathcal{D}_{r,l}$
  - Disjoint cross-view, unlabeled dataset $\mathcal{D}_u$
- Inference:
  - Input from target view, v_t
  - Predict on both reference, and target views



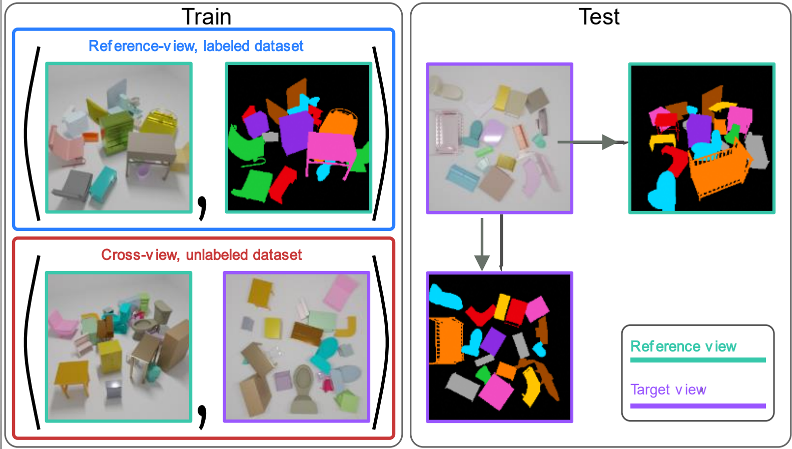
Zero pair, cross-view semantic segmentation



- New semi-supervised task:

  Zero-Pair, Cross-View semantic segmentation

- Train on:
  - Reference view (v_r), labeled dataset $\mathcal{D}_{r,l}$
  - Disjoint cross-view, unlabeled dataset $\mathcal{D}_u$
- Inference:
  - Input from target view, v_t
  - Predict on both reference, and target views



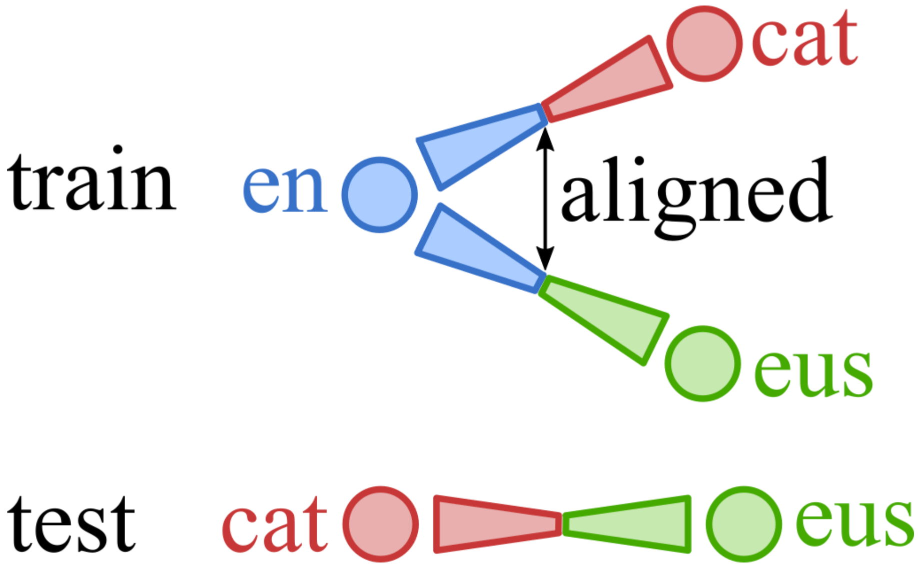
6-ZPCVNet

Related work

Mix&Match Networks [Wang2020]

Translation between domain/modality pairs not seen during training

Enforcing latent space alignment of encoder-decoder pairs

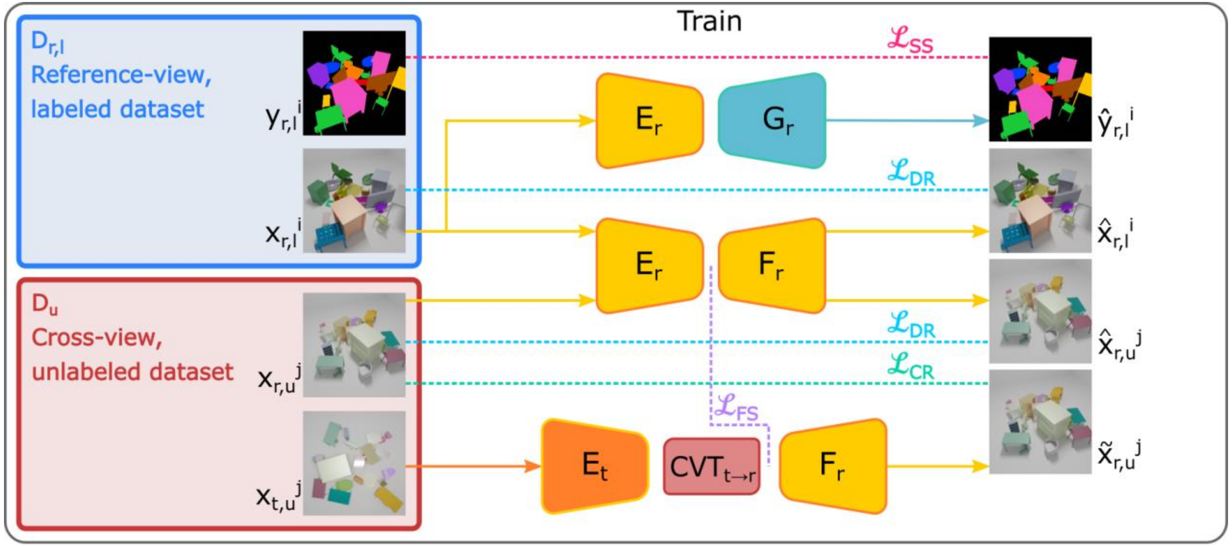


[Johnson2016] M. Johnson et al., "Google's multilingual neural machine translation system: enabling zero-shot translation," arXiv, 2016

[Wang2020] Y. Wang et al., "Mix and match networks: multi-domain alignment for unpaired image-to-image translation," *IJCV* 2020

Model

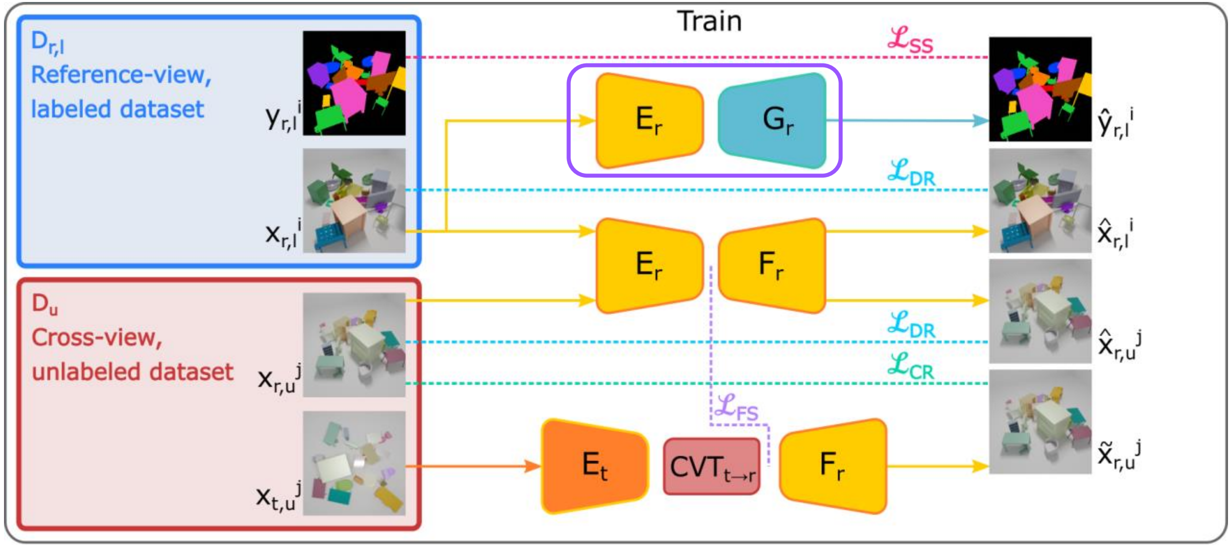
ZPCVNet



- ZPCVNet modules:
  - Reference-view fully-supervised semantic segmentation Encoder-Decoder [Er-Gr]
  - Reference-view Autoencoder [Er-Fr]
  - Cross-view Encoder-Decoder on RGB views [Et-CVT-Fr]
- Seek latent space alignment
  - Cross-View Transformer
  - Shared weights
  - Pseudolabels

Model

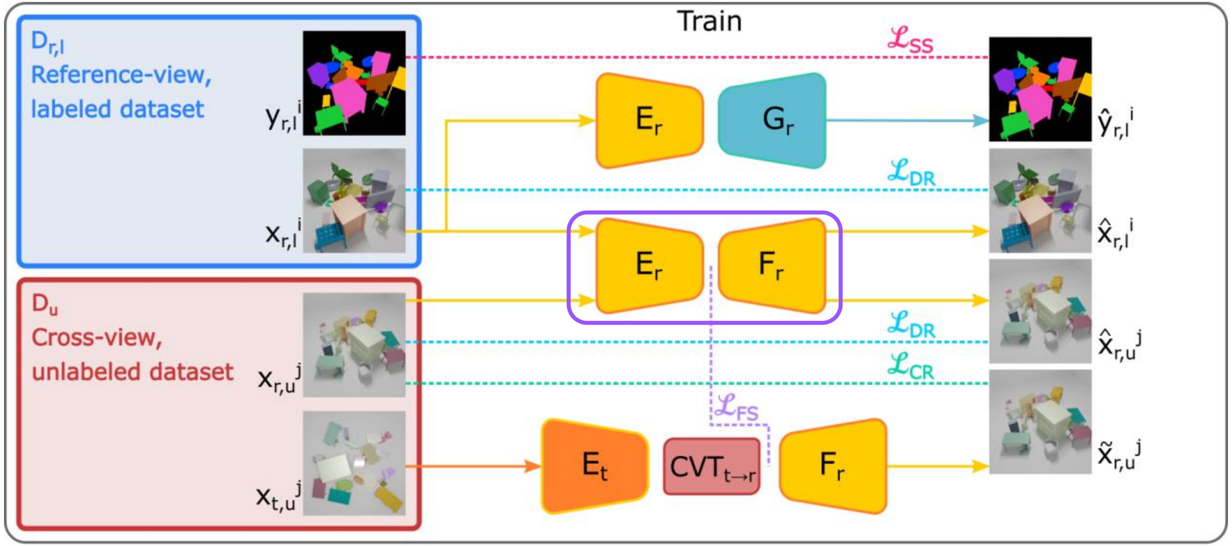
ZPCVNet



- ZPCVNet modules:
  - Reference-view fully-supervised semantic segmentation Encoder-Decoder [Er-Gr]
  - Reference-view Autoencoder [Er-Fr]
  - Cross-view Encoder-Decoder on RGB views [Et-CVT-Fr]
- Seek latent space alignment
  - Cross-View Transformer
  - Shared weights
  - Pseudolabels

Model

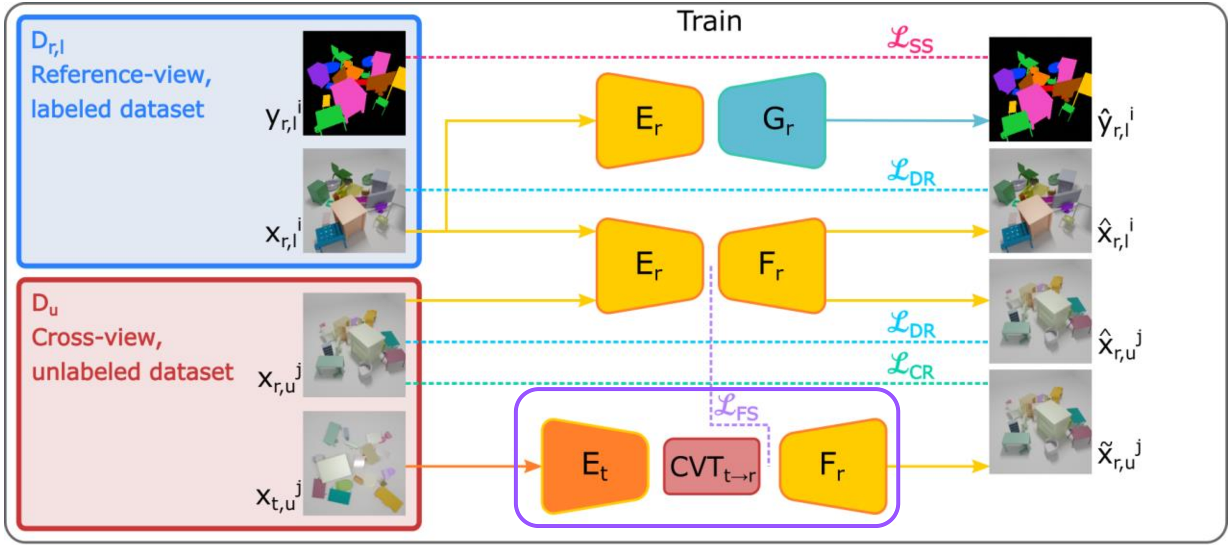
ZPCVNet



- ZPCVNet modules:
  - Reference-view fully-supervised semantic segmentation Encoder-Decoder [Er-Gr]
  - Reference-view Autoencoder [Er-Fr]
  - Cross-view Encoder-Decoder on RGB views [Et-CVT-Fr]
- Seek latent space alignment
  - Cross-View Transformer
  - Shared weights
  - Pseudolabels

Model

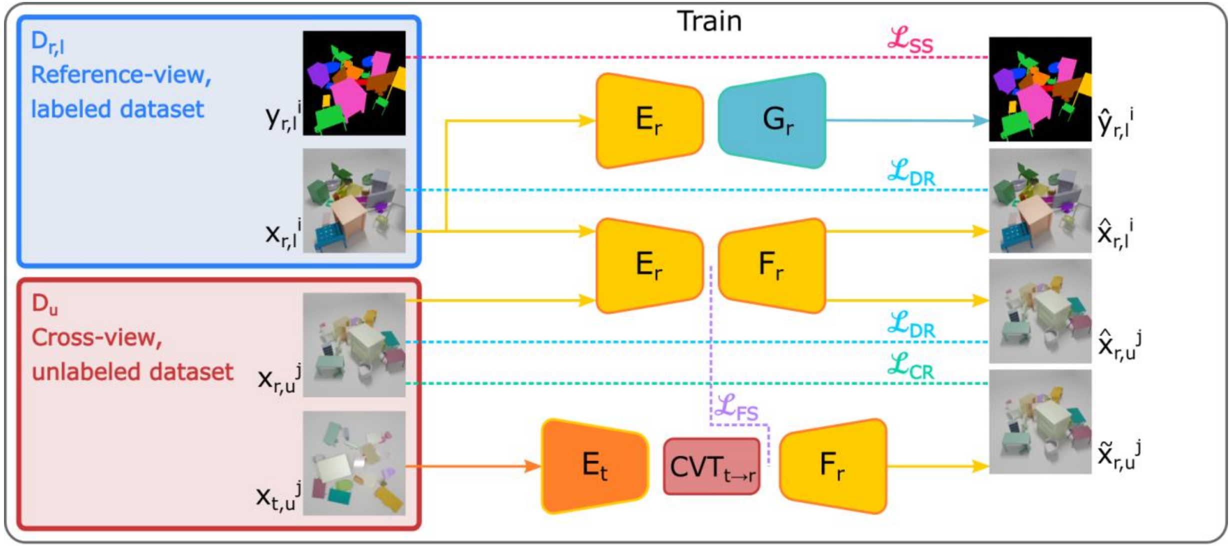
ZPCVNet



- ZPCVNet modules:
  - Reference-view fully-supervised semantic segmentation Encoder-Decoder [Er-Gr]
  - Reference-view Autoencoder [Er-Fr]
  - Cross-view Encoder-Decoder on RGB views [Et-CVT-Fr]
- Seek latent space alignment
  - Cross-View Transformer
  - Shared weights
  - Pseudolabels

Model

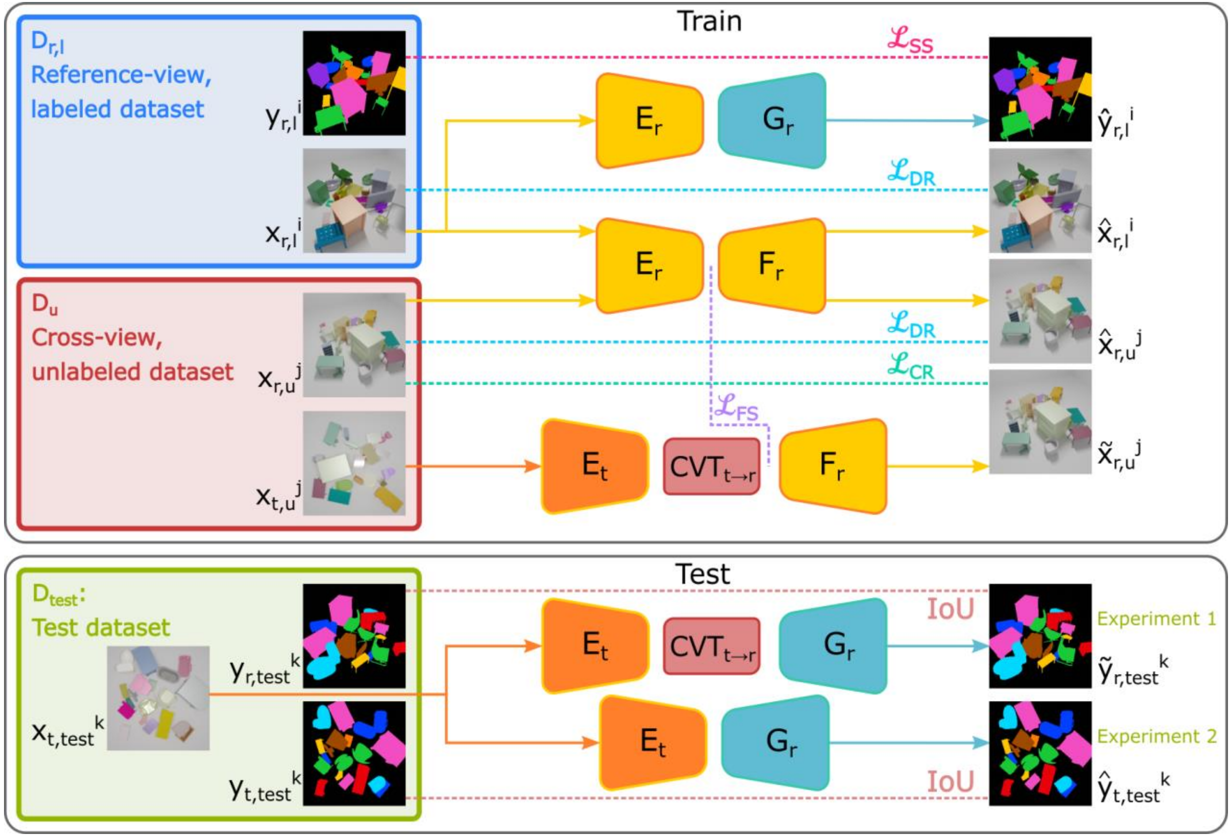
ZPCVNet



- ZPCVNet modules:
  - Reference-view fully-supervised semantic segmentation Encoder-Decoder [Er-Gr]
  - Reference-view Autoencoder [Er-Fr]
  - Cross-view Encoder-Decoder on RGB views [Et-CVT-Fr]
- Seek latent space alignment
  - Cross-View Transformer
  - Shared weights
  - Pseudolabels

Model

ZPCVNet

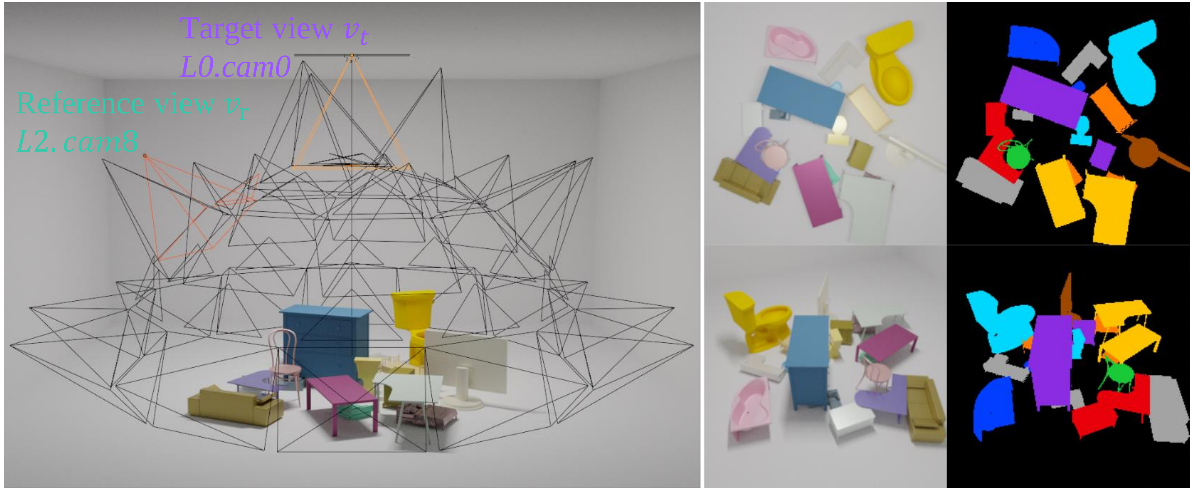


- ZPCVNet modules:
  - Reference-view fully-supervised semantic segmentation Encoder-Decoder [Er-Gr]
  - Reference-view Autoencoder [Er-Fr]
  - Cross-view Encoder-Decoder on RGB views [Et-CVT-Fr]
- Seek latent space alignment
  - Cross-View Transformer
  - Shared weights
  - Pseudolabels
- Inference: Semantic predictions on viewpoints that have no semantic ground truth

Experiments

Setup

Dataset: MVMO (*Other Objects* subset, 64x64)



Train with no available ground truth from the new viewpoint. Fully-supervised approach is not possible.

½ batch from each dataset (labeled $\mathcal{D}_{r,l}$, unlabeled $\mathcal{D}_u$)

Experiment 1

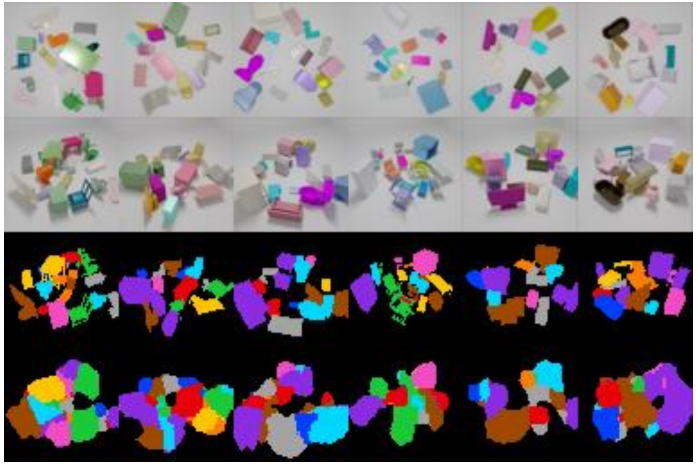
Cross-view with output in v_r

Input (v_t)

v_r (not provided)

Ground truth (in v_r)

Ours



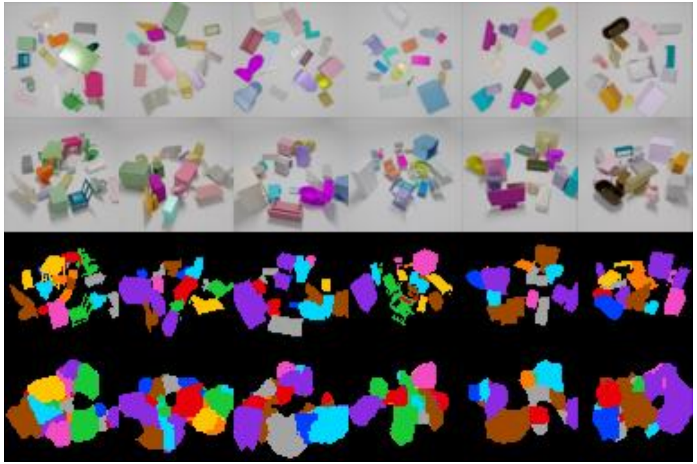
Experiment1: Output in v_r

Method	Trained on	bathtub	bed	chair	desk	dresser	monitor	nightstand	sofa	table	toilet	noBG	IoU Avg.
FSCV	$x^i_{t,l}, y^i_{r,l} \sim \mathcal{D}^+_{r,l}$	04.81	03.54	04.43	00.00	00.00	00.00	05.64	00.00	04.43	00.00	02.29	07.45
$\text{CV}_{test}v_r$	$x^i_{r,l}, y^i_{r,l} \sim \mathcal{D}_{r,l}$	02.53	03.28	00.82	01.14	01.27	02.42	00.59	02.92	01.56	00.55	01.71	07.04
$\text{CV}_{test}v_t$	$x^i_{t,l}, y^i_{t,l} \sim \mathcal{D}^+_{r,l}$	02.22	02.65	02.82	03.11	03.28	02.58	03.58	02.47	02.08	02.81	02.76	07.81
Homography	$x^i_{r,l}, y^i_{r,l} \sim \mathcal{D}_{r,l}$	06.25	09.48	01.22	03.14	01.68	04.59	01.03	07.76	03.79	01.72	04.06	10.65
Mix&Match	$x^i_{r,l}, y^i_{r,l} \sim \mathcal{D}_{r,l}$ $x^j_{r,u}, x^j_{t,u} \sim \mathcal{D}_u$	02.04	03.28	03.47	04.32	02.46	03.84	01.63	03.00	02.57	03.95	3.06	08.33
ZPCVNet (ours)	$x^i_{r,l}, y^i_{r,l} \sim \mathcal{D}_{r,l}$ $x^j_{r,u}, x^j_{t,u} \sim \mathcal{D}_u$	**17.73**	**21.47**	**14.33**	**10.81**	**18.52**	**22.82**	**22.73**	**12.81**	**14.29**	**20.42**	**17.59**	**23.10**
												Upper bound: train-test on v_r	33.43

6-ZPCVNet

Experiment 1

Cross-view with output in v_r



Experiment1: Output in v_r

Method	CVT	PL	noBG	Avg.
Vanilla	No	No	03.17	09.26
Vanilla+PL	No	Yes	02.80	08.70
Vanilla+CVT	Yes	No	05.73	11.41
ZPCVNet (full model)	Yes	Yes	**17.59**	**23.10**

Ablation: both CVT and pseudolabels are needed

Method	Trained on	bathtub	bed	chair	desk	dresser	monitor	nightstand	sofa	table	toilet	noBG	IoU Avg.
FSCV	$x^i_{t,l}, y^i_{r,l} \sim \mathcal{D}^+_{r,l}$	04.81	03.54	04.43	00.00	00.00	00.00	05.64	00.00	04.43	00.00	02.29	07.45
$\mathrm{CV}_{test} v_r$	$x^i_{r,l}, y^i_{r,l} \sim \mathcal{D}_{r,l}$	02.53	03.28	00.82	01.14	01.27	02.42	00.59	02.92	01.56	00.55	01.71	07.04
$\mathrm{CV}_{test} v_t$	$x^i_{t,l}, y^i_{t,l} \sim \mathcal{D}^+_{r,l}$	02.22	02.65	02.82	03.11	03.28	02.58	03.58	02.47	02.08	02.81	02.76	07.81
Homography	$x^i_{r,l}, y^i_{r,l} \sim \mathcal{D}_{r,l}$	06.25	09.48	01.22	03.14	01.68	04.59	01.03	07.76	03.79	01.72	04.06	10.65
Mix&Match	$x^i_{r,l}, y^i_{r,l} \sim \mathcal{D}_{r,l}$; $x^j_{r,u}, x^j_{t,u} \sim \mathcal{D}_u$	02.04	03.28	03.47	04.32	02.46	03.84	01.63	03.00	02.57	03.95	3.06	08.33
ZPCVNet (ours)	$x^i_{r,l}, y^i_{r,l} \sim \mathcal{D}_{r,l}$; $x^j_{r,u}, x^j_{t,u} \sim \mathcal{D}_u$	**17.73**	**21.47**	**14.33**	**10.81**	**18.52**	**22.82**	**22.73**	**12.81**	**14.29**	**20.42**	**17.59**	**23.10**
Upper bound: train-test on v_r													33.43

Experiment 2

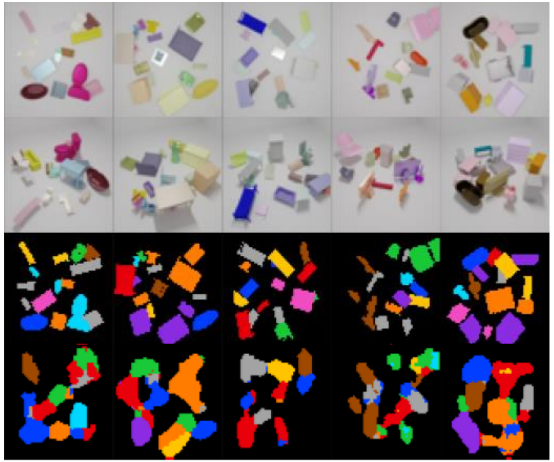
Cross-view with output in v_t

Input (v_t)

v_r (not provided)

Ground truth

Ours



Experiment2: Output in v_t

Method	Trained on	bathtub	bed	chair	desk	dresser	monitor	nightstand	sofa	table	toilet	noBG	IoU Avg.
FSv_t ⋆(upper bound)	$x^i_{t,l}, y^i_{t,l} \sim \mathcal{D}^+_{r,l}$	30.86	33.11	24.70	18.88	25.19	29.39	33.06	22.14	11.80	32.35	26.15	31.67
CV$_{test}$	$x^i_{r,l}, y^i_{r,l} \sim \mathcal{D}_{r,l}$	09.87	10.50	03.39	04.67	03.67	09.16	03.15	10.91	05.96	04.31	06.56	13.29
Homography	$x^i_{r,l}, y^i_{r,l} \sim \mathcal{D}_{r,l}$	06.49	08.49	01.12	03.10	02.03	05.11	01.46	06.96	05.97	02.35	04.31	10.42
ZPCVNet (ours)	$x^i_{r,l}, y^i_{r,l} \sim \mathcal{D}_{r,l}$ $x^j_{r,u}, x^j_{t,u} \sim \mathcal{D}_u$	**16.40**	**17.07**	**09.57**	**10.13**	**06.54**	**14.87**	**06.73**	**14.95**	**11.20**	**11.80**	**11.93**	**18.10**

Takeaways

- New semi-supervised task: ***Zero-Pair, Cross-View semantic segmentation***
- ZPCVNet model:
  - Reasonable predictions on both references with one model (pluging CVT in/out)
  - Outperforms other learned (deep) and geometry-based baselines over MVMO
  - Initial baseline for further research in dense semantic knowledge transfer across views

Publication **A. Alvarez-Gila**, J. Van De Weijer, Y. Wang, and E. Garrote, "*Zero-Pair Semi-Supervised Cross-View Semantic Segmentation,*" 3DV 2022 (under review).

Self-supervised learning for image-to-image translation in the small data regime

Conclusions

2 Self-Supervised Blur Detection from Synthetically Blurred Scenes

Contribution: Learning framework for self/weak/semi-supervised blur detection from synthetic degradation model. State-of-the art results without access to real blur masks.

3 Adversarial Networks for Spatial Context-Aware Spectral Image Reconstruction from RGB

Contribution: Method for spectral super-resolution from RGB images leveraging spatial context, achieving state-of-the art results.

4 A Probabilistic Model and Capturing Device for Remote Simultaneous Estimation of Spectral Emissivity and Temperature of Hot Emissive Materials

Contribution: Device and general Bayesian probabilistic method achieving remote online estimates for temperature and spectral emissivity with quantification of uncertainty, with a real application in EAF-based steelmaking.

Conclusions

5 MVMO: A Multi-Object Dataset for Wide Baseline Multi-View Semantic Segmentation

Contribution: New synthetic dataset (and code) enabling research in novel research lines: multi-view and cross-view semantic segmentation.

Future work:

(i) Code extension for additional modalities to support multi-modal and object-centric tasks.

6 Zero-Pair Semi-Supervised Cross-View Semantic Segmentation

Contribution: Novel semi-supervised task of zero-pair, cross-view semantic segmentation applicable to inline industrial scenarios. New method (ZPCVNet) outperforming geometric and deep baselines.

Future work:

(i) Adaptation of 2nd wave self-supervised methods to multi/cross view dense prediction tasks.

(ii) Inclusion of inductive biases from multiple-view geometry (e.g. epipolar constraint)

(iii) Attention-guided Cross View Transformer for dynamic spatial feature mapping across views.

Summary of published works (I)

SynthBlur [1] **A. Alvarez-Gila**, A. Galdran, E. Garrote, and J. van de Weijer, "*Self-supervised blur detection from synthetically blurred scenes,*" Image and Vision Computing, 2019.

RGB2HSI [2] **A. Alvarez-Gila**, J. van de Weijer, and E. Garrote, "*Adversarial Networks for Spatial Context-Aware Spectral Image Reconstruction from RGB,*" ICCVW 2017

RGB2HSI [3] B. Arad, O. Ben-Shahar, R. Timofte, L. Van Gool, L. Zhang, M.-H. Yang, Z. Xiong, C. Chen, Z. Shi, D. Liu, F. Wu, C. Lanaras, S. Galliani, K. Schindler, T. Stiebel, S. Koppers, P. Seltsam, R. Zhou, M. El Helou, F. Lahoud, M. Shahpaski, K. Zheng, L. Gao, B. Zhang, X. Cui, H. Yu, Y. B. Can, **A. Alvarez-Gila**, J. van de Weijer, E. Garrote, A. Galdran, M. Sharma, S. Koundinya, A. Upadhyay, R. Manekar, R. Mukhopadhyay, H. Sharma, S. Chaudhury, K. Nagasubramanian, S. Ghosal, A. K. Singh, A. Singh, B. Ganapathysubramanian, and S. Sarkar, "*NTIRE 2018 Challenge on Spectral Reconstruction from RGB Images,*" CVPRW 2018

TES [4] A. Picon*, **A. Alvarez-Gila***, J. A. Arteche, G. A. López, and A. Vicente, "*A Probabilistic Model and Capturing Device for Remote Simultaneous Estimation of Spectral Emissivity and Temperature of Hot Emissive Materials,*" IEEE Access, 2021. ***Equal contribution**

TES [5] A. Picon, **A. Alvarez-Gila**, A. Vicente, and Arteche, Jose Antonio, "*System and method for determining the emitting temperature and emissivity in a wavelength range of metallurgical products,*" PCT/IB2019/061335 , filed December 24, 2019

MVMO [6] **A. Alvarez-Gila**, J. Van De Weijer, Y. Wang, and E. Garrote, "*MVMO: A Multi-Object Dataset for Wide Baseline Multi-View Semantic Segmentation,*" ICIP 2022.

ZPCVNet [7] **A. Alvarez-Gila**, J. Van De Weijer, Y. Wang, and E. Garrote, "*Zero-Pair Semi-Supervised Cross-View Semantic Segmentation,*" 3DV 2022 (under review).

Summary of published works (II)

Side projects at Tecnalia - Publications

Inverse problems – dehazing:

[8] A. Galdran, **A. Alvarez-Gila**, A. Bria, J. Vazquez-Corral, and M. Bertalmío, "On the Duality Between Retinex and Image Dehazing," CVPR 2018.

[9] C. Ancuti, C. O. Ancuti, R. Timofte, L. Van Gool, L. Zhang, M.-H. Yang, V. M. Patel, H. Zhang, V. A. Sindagi, R. Zhao, X. Ma, Y. Qin, L. Jia, K. Friedel, S. Ki, H. Sim, J.-S. Choi, S. Kim, S. Seo, S. Kim, M. Kim, R. Mondal, S. Santra, B. Chanda, J. Liu, K. Mei, J. Li, Luyao, F. Fang, A. Jiang, X. Qu, T. Liu, P. Wang, B. Sun, J. Deng, Y. Zhao, M. Hong, J. Huang, Y. Chen, E. Chen, X. Yu, T. Wu, A. Genc, D. Engin, H. K. Ekenel, W. Liu, T. Tong, G. Li, Q. Gao, Z. Li, D. Tang, Y. Chen, Z. Huo, **A. Alvarez-Gila**, A. Galdran, A. Bria, J. Vazquez-Corral, M. Bertalmío, H. S. Demir, O. F. Adil, H. X. Phung, X. Jin, J. Chen, C. Shan, and Z. Chen, "*NTIRE 2018 Challenge on Image Dehazing: Methods and Results,*" CVPRW 2018.

Agricultural image understanding:

[10] D. Argüeso, A. Picon, U. Irusta, A. Medela, M. G. San-Emeterio, A. Bereciartua, and **A. Alvarez-Gila**, "Few-Shot Learning approach for plant disease classification using images taken in the field," Computers and Electronics in Agriculture, 2020.

[11] A. Picon, M. Seitz, **A. Alvarez-Gila**, P. Mohnke, A. Ortiz-Barredo, and J. Echazarra, "Crop conditional Convolutional Neural Networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions," Computers and Electronics in Agriculture, 2020.

[12] A. Picon, **A. Alvarez-Gila**, M. Seitz, A. Ortiz-Barredo, J. Echazarra, and A. Johannes, "Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild," Computers and Electronics in Agriculture, 2019.

[13] A. Johannes, A. Picon, **A. Alvarez-Gila**, J. Echazarra, S. Rodriguez-Vaamonde, A. D. Navajas, and A. Ortiz-Barredo, "Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case," Computers and Electronics in Agriculture, 2017.

Medical imaging/signal analysis:

[14] A. Picon, U. Irusta, **A. Alvarez-Gila**, E. Aramendi, F. Alonso-Atienza, C. Figuera, U. Ayala, E. Garrote, L. Wik, J. Kramer-Johansen, and T. Eftestøl, "Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia," PLOS ONE, 2019.

[15] A. Picon, U. Irusta, **A. Alvarez-Gila**, E. Aramendi, E. Garrote, U. Ayala, F. Alonso, and C. Figuera, "Detección de fibrilación ventricular mediante técnicas de aprendizaje profundo," CASEIB 2017.

[16] A. Galdran, **A. Alvarez-Gila**, M. I. Meyer, C. L. Saratxaga, T. Araújo, E. Garrote, G. Aresta, P. Costa, A. M. Mendonça, and A. Campilho, "Data-Driven Color Augmentation Techniques for Deep Skin Image Analysis," arXiv:1703.03702 [cs], 2017.

Industry:

[17] A. Picon Ruiz, **A. Alvarez-Gila**, U. Irusta, and J. Echazarra Huguet, "Why deep learning performs better than classical machine learning approaches," DYNA, 2020.

Summary of published works (III)

Side projects at Tecnalia - Patents

Agricultural image understanding:

[18] D. Roldan Lopez, J. Romero Rodriguez, C. M. Spangler, C. Klukas, T. Eggers, R. Navarra-mestre, A. M. Ortiz Barredo, **A. Alvarez-Gila**, J. Echazarra Huguet, A. Picon Ruiz, And A. Bereciartua Perez, “Quantifying plant infestation by estimating the number of insects on leaves, by convolutional neural networks that provide density maps,” EP 19200657.5 (191150EP01).

[19] A. Picon, M. Nachtmann, M. Seitz, P. Mohnke, R. Navarra-Mestre, A. Johannes, T. Eggers, A. O. Barredo, **A. Alvarez-Gila**, and J. Echazarra, “System and Method for Plant Disease Detection Support,” 21-May-2019.

[20] A. Johannes, T. Eggers, A. Picon, **A. Alvarez-Gila**, A. M. Ortiz Barredo, and A. M. Díez-Navajas, “System and Method for Detecting Plant Diseases,” WO/2017/194276, 17-Nov-2017.

Medical imaging/signal analysis:

[21] E. Garrote, A. Bereciartua, A. Picon, C. Lopez-Saratxaga, **A. Alvarez-Gila**, A. Galdran, R. Bilbao, and O. Belar, “Analysing histological images,” PCT/ES2016/070881.

Summary of published code/data

SynthBlur, results and info on the method presented in Chapter 2 and in "*Self-supervised blur detection from synthetically blurred scenes,*" Image and Vision Computing, 2019,
https://github.com/aitorshuffle/synthblur

adv_rgb2hsi, models and test scripts corresponding to our submission to the NTIRE 2018 NTIRE 2018 Challenge on Spectral Reconstruction from RGB Images," CVPRW 2018
https://github.com/aitorshuffle/ntire2018_adv_rgb2hs

MVMO, code and dataset (soon) presented in Chapter 5 and in *MVMO: A Multi-Object Dataset for Wide Baseline Multi-View Semantic Segmentation*," ICIP 2022,
https://aitorshuffle.github.io/projects/mvmo/

Self-supervised learning for image-to-image translation in the small data regime

PhD thesis dissertation by **Aitor Álvarez Gila**

Directors:

Joost van de Weijer

Estibaliz Garrote

CVC, Bellaterra

2022-07-19

Universitat Autònoma de Barcelona

Computer Vision Center

MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE