Instituto Universitario de Investigación Ingeniería de Aragón Universidad Zaragoza 1

PanoRoom: From the Sphere to the 3D Layout

Laboratoire Électronique Informatique et Image 2

Clara Fernández-Labrador[1,2], José M. Fácil[1], Alejandro Pérez-Yus[1], Cédric Demonceaux[2], José J. Guerrero[1]

Overview

Motivation

Scan the **QR Code** to see our paper!

Related Work

[1] Fernandez-Labrador, C., Perez-Yus, A., Lopez-Nicolas, G., Guerrero, J.J.: Layouts from panoramic images with geometry and deep learning. RA-L/IROS 2018.

[2] C. Zou, A. Colburn, Q. Shan and D. Hoiem. "Layoutnet: Reconstructing the 3D room layout from a single rgb image". CVPR 2018.

[3] Y. Zhang, S. Song, P. Tan, and J. Xiao. "PanoContext: A whole-room 3D context model for panoramic scene understanding." ECCV 2014.

Acknowledgements

This work was supported by Projects DPI2014-61792-EXP, DPI2015-67275 (MINECO/FEDER), DPI2015-65962-R (UE) and by the Regional Council of Burgundy (CRB).

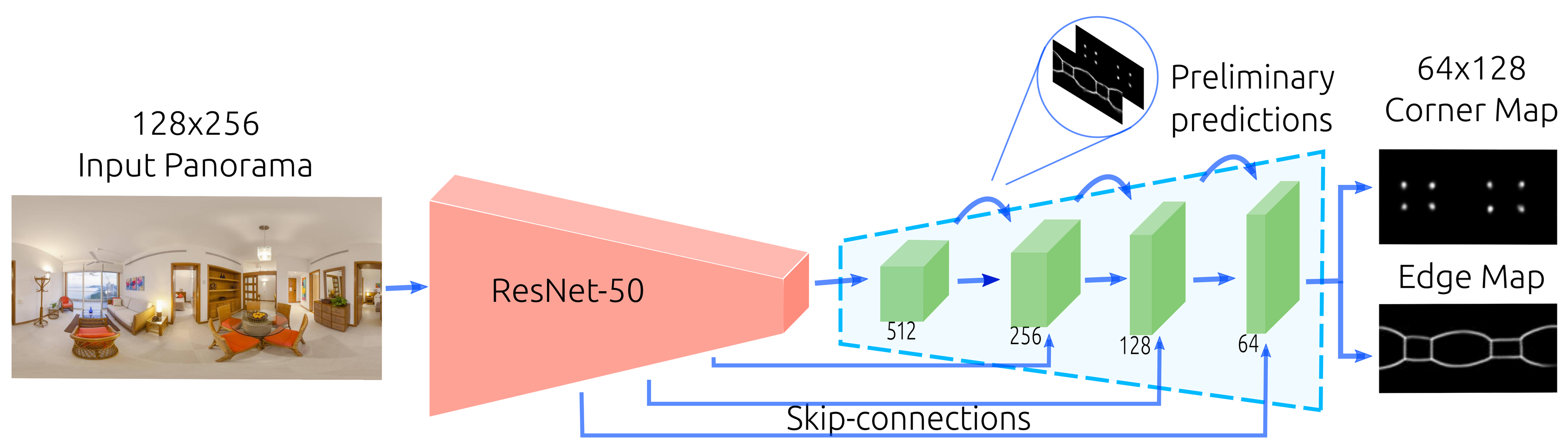
Encoder-Decoder FCN Details

Encoder

We build our model over **ResNet-50** pre-trained on **ImageNet** dataset.
- Residual networks allow to increase depth without increasing the number of parameters with respect to their plain counterparts.
- Faster convergence due to the general low-level features learned from ImageNet.

Decoder

- We propose a **unique branch** for multi-tasking to reduce computation time and the number of parameters.
- We introduce **skip-connections** between encoder and decoder.
- We perform **preliminary predictions** in different resolutions which are concatenated.



Loss Functions

Pixel-wise sigmoid cross-entropy loss

These maps have an extremely unbalanced distribution of edges and corners, so we introduce the **ponder factors λ_1 and λ_0.**

$$\mathcal{L}_{Mi} = \lambda_1 (y_i(-\log(S(\hat{y}_i)))) + \lambda_0((1-y_i)(-\log(1-S(\hat{y}_i))))$$

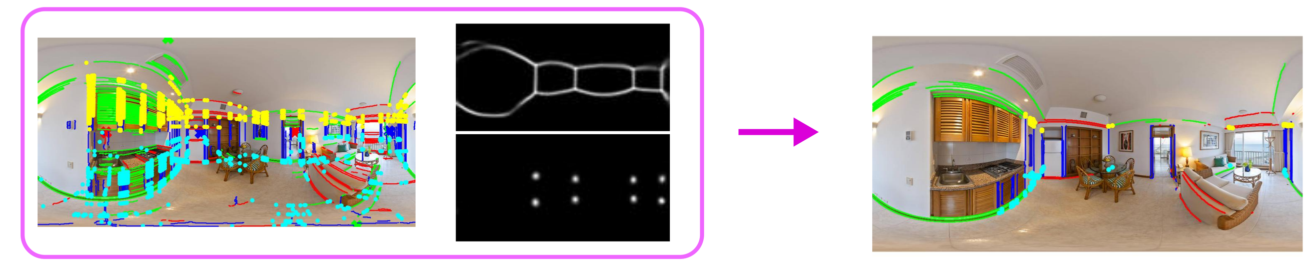
Perceptual loss

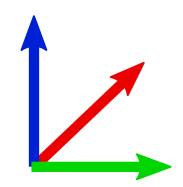
Apart from encouraging the output images $\hat{y}$ to match the target images y, we also encourage them to have **same feature representations.**

$$\mathcal{L}_P(y,\mathcal{I}) = ||\hat{\phi}_j(\mathcal{I}) - \phi_j(y)||_2^2$$

Layout Recovery and Results

Layout Hypotheses Generation and Evaluation





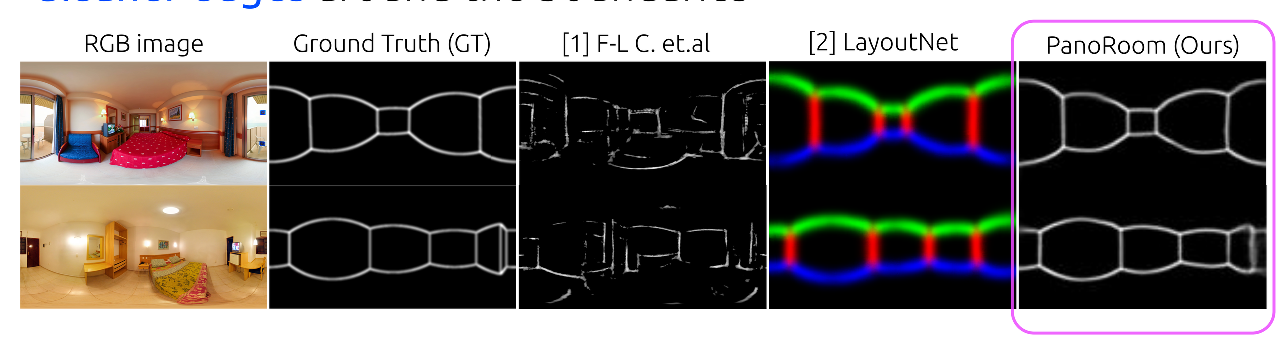
We randomly sample corners to generate layout hypotheses following the Manhattan World Assumption. **We include the possibility of introducing new corners** for those cases that they are not visible due to the scene convexity or occlusions.

Finally we choose the best fitting solution with the **Edge and Corners Maps** obtained trough our FCN by the following function:

$$L^{Best} = \text{argmax}(w_e \sum P_{edge}(L_E^h) + w_c \sum P_{corner}(L_C^h))$$

FCN Evaluation

- **96.2%** accuracy on SUN360 dataset.
- **Able to generalize**: 91.6% acc. on Stanford2D-3D dataset.
- **General Layouts**: Not limited to box-type
- **Cleaner edges** around the boundaries



3D Layout Evaluation

We **outperform the State of the Art** in all the metrics while being faster.

Dataset	Method	3DIoU(%)	CE(%)	PE^{SS}(%)	PE^{CS}(%)
SUN360	PanoContext [3]	67.22	1.60	4.55	10.34
	F-L C. *et al.* [1]	-	-	-	7.26
	LayoutNet [2]	74.48	1.06	3.34	-
	Ours	**76.82**	**0.79**	**2.59**	**3.13**
Stnfd.2D-3D	F-L C. *et al.* [1]	-	-	-	12.1
	Ours	**70.64**	**1.15**	**3.95**	**4.98**

Method	Computation Time (s)
PanoContext [19]	>300
LayoutNet [22]	44.73
Ours	10.07

Input Image with Reprojected 2D Layout | 3D Model