



Desde imágenes RGB a mapas de profundidad pasando por redes convolucionales profundas

Zuria Bauer Hartwig

Departamento de Ciencia de la Computación e Inteligencia Artificial

13 de diciembre de 2018

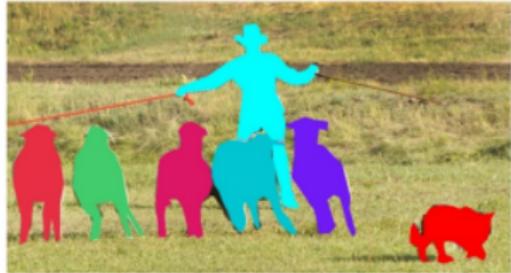
RoViT



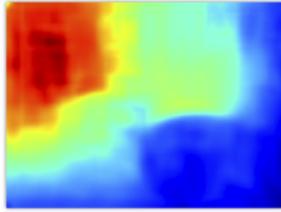


- PhD en Ciencias de la Computación en Inteligencia Artificial (actual)
- M.Sc.Automática y Robótica
- Grado en Ingeniería Química
- Programando con Python desde 2014
- Ponente en la PyConES 2014 en Zaragoza, EuroPython 2016 en Bilbao y en PyConES 2016 en Almería x2y 2018 en Málaga
- Miembro activo de [CACHemE.org](#) y [Python Alicante](#)
- Idiomas: Español, Inglés y Alemán
- [@zuriich92](#)
- zbauer@dccia.ua.es

Diversos problemas de visión por computación se han visto beneficiados por la incorporación de la información de profundidad.



Single RGB Image



Depth Map

A "working" Monocular Depth System

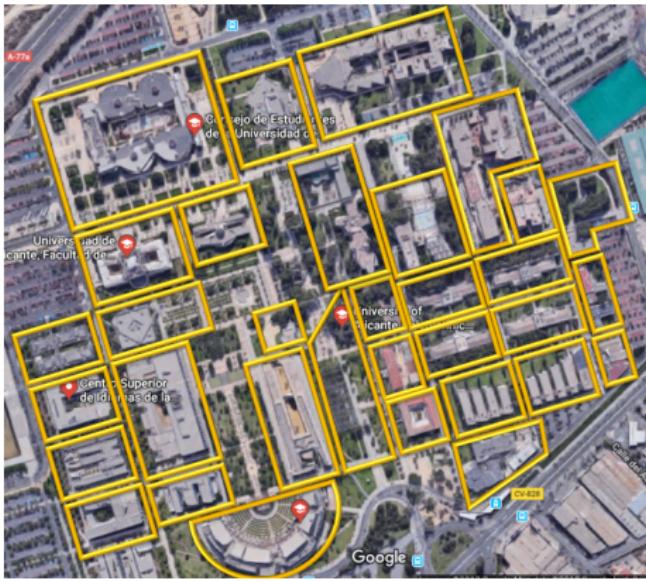


Universitat d'Alacant
Universidad de Alicante

Universidad de Alicante



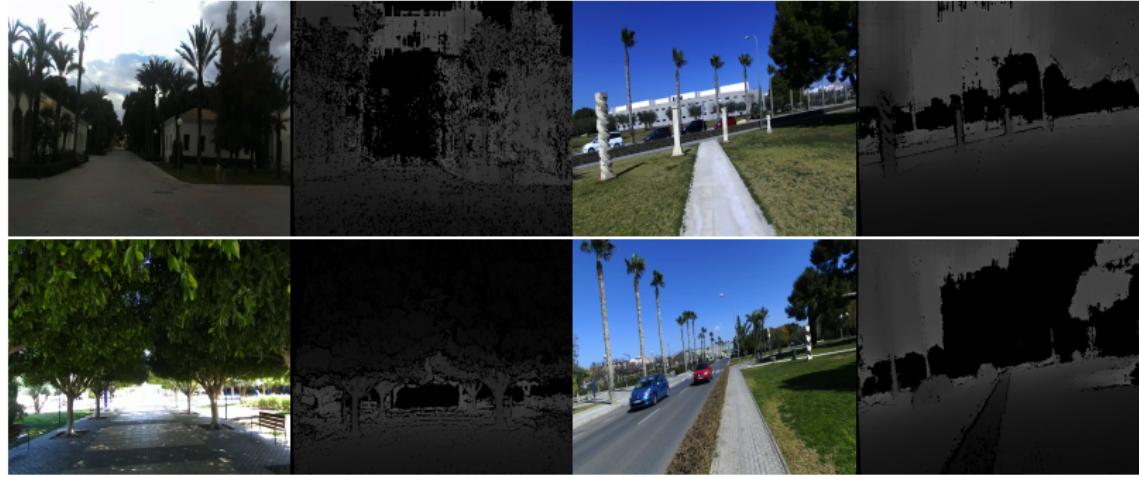
Dataset	Scale			Modality			
	Frames	Layouts	Resolution	RGB	Depth	Point Cloud	Outdoor
SCENENET RGB-D []	5M	57	320x240	✓	✓	✗	✗
MATTERPORT 3D []	200k	90	1280x1024	✓	✓	✓	✗
STANFORD 2D-3D []	70k	270	1080x1080	✓	✓	✓	✗
SCENE NN []	2k-10k each scene	100	640x480	✓	✓	✗	✗
MICROSOFT RGB-D []	500-1k each seq.	7	640x480	✓	✓	✗	✗
VIDRILLO []	22k	10	640x480	✓	✗	✓	✗
SUN RGBD []	10k	270	1080x1080	✓	✓	✓	✗
NYU-D V2 []	1.5k	464	640x480	✓	✓	✗	✗
B3DO []	849	75	1080x1080	✓	✓	✗	✗
SYNTHIA []	200k		960x720	✓	✓	✗	✓
KITTI []	1.6k	400	1392 x 512	✓	✓	✓	✓
ETH3D []	898	25		✓	✓	✗	✓
MAKE 3D []	534		1280x1024	✓	✓	✗	✓
OURS	482736	33	2208x1242	✓	✓	✗	✓

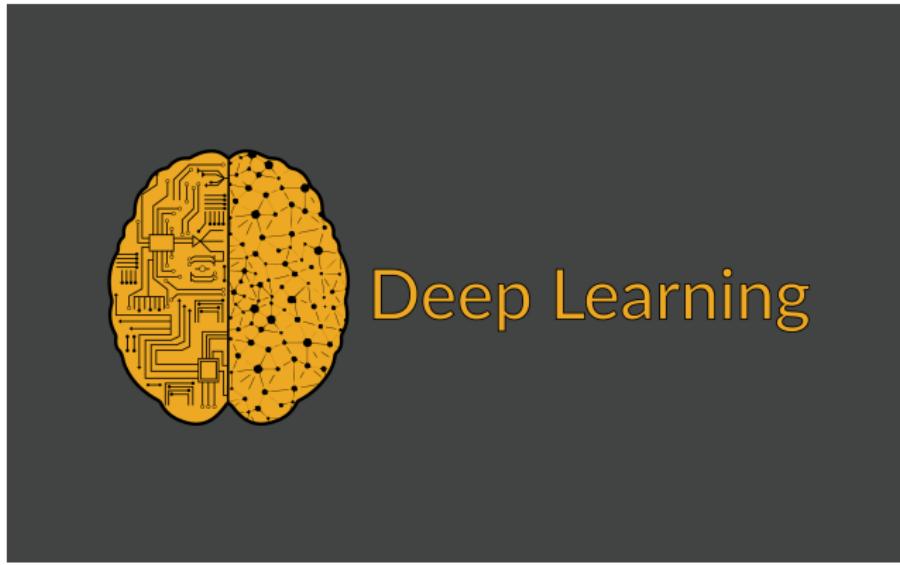


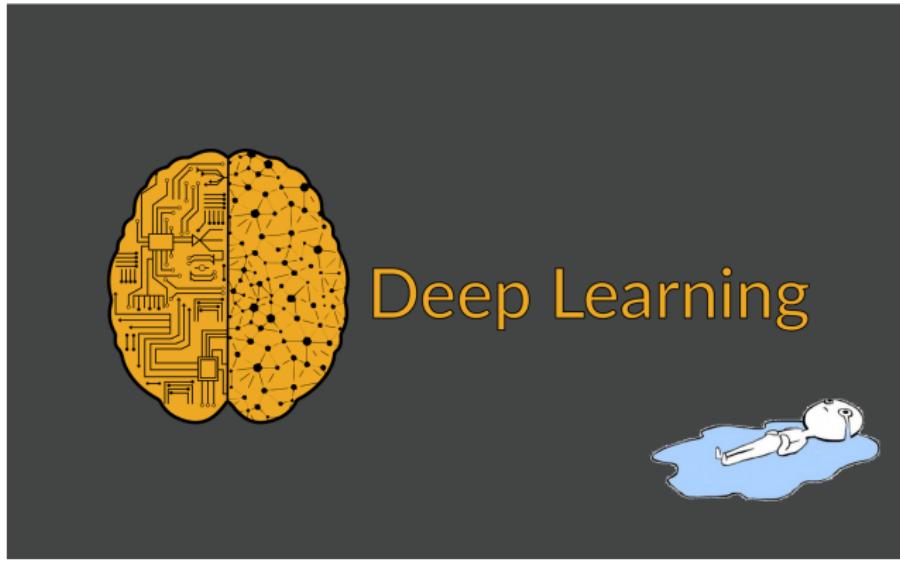
- **UASOL:** Large-scale High-resolution Stereo Outdoor Dataset.
- Secuencias Completas de pares Stereo con sus correspondientes mapas de profundidad.
- Resolución: 2208×1242 .
- Grabado a 15 frames por segundo.
- Point of view of a pedestrian.
- Características visuales diversas introduciendo condiciones meteorológicas cambiantes, distintos tipos de iluminación, peatones, vehículos, otros medios de transporte, etc.
- **482736** frames distribuidos en 33 secuencias.
- Tracking posicional de cada frame.
- Parámetros de calibración de la cámara para cada escena.

ZED







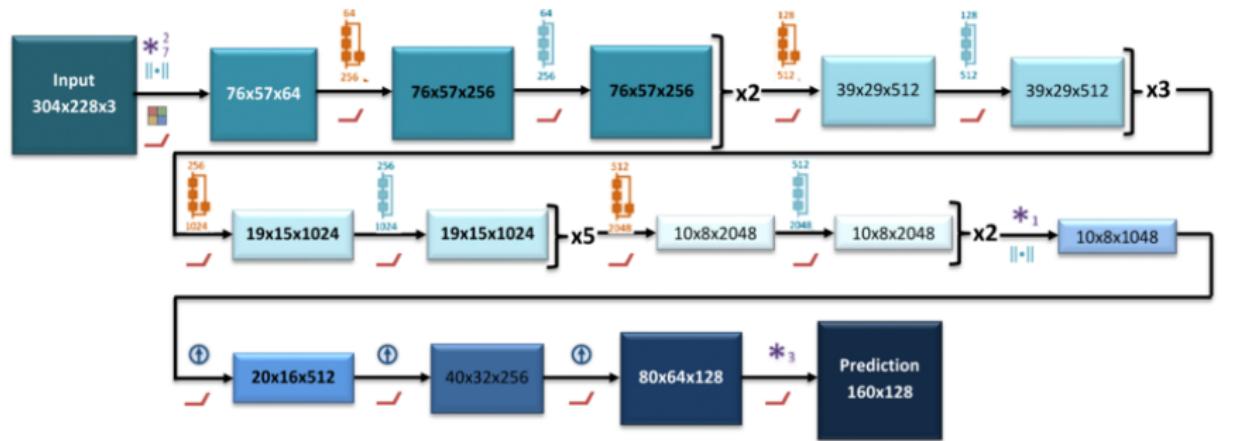


Depth from single monocular images

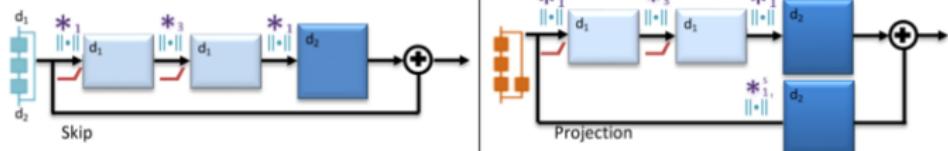


Universitat d'Alacant
Universidad de Alicante

Universidad de Alicante



- \ast_n : Nxn convolution (stride s)
- ReLU
- 3x3 max-pooling, stride 2
- Batch Normalization
- Up- Convolution

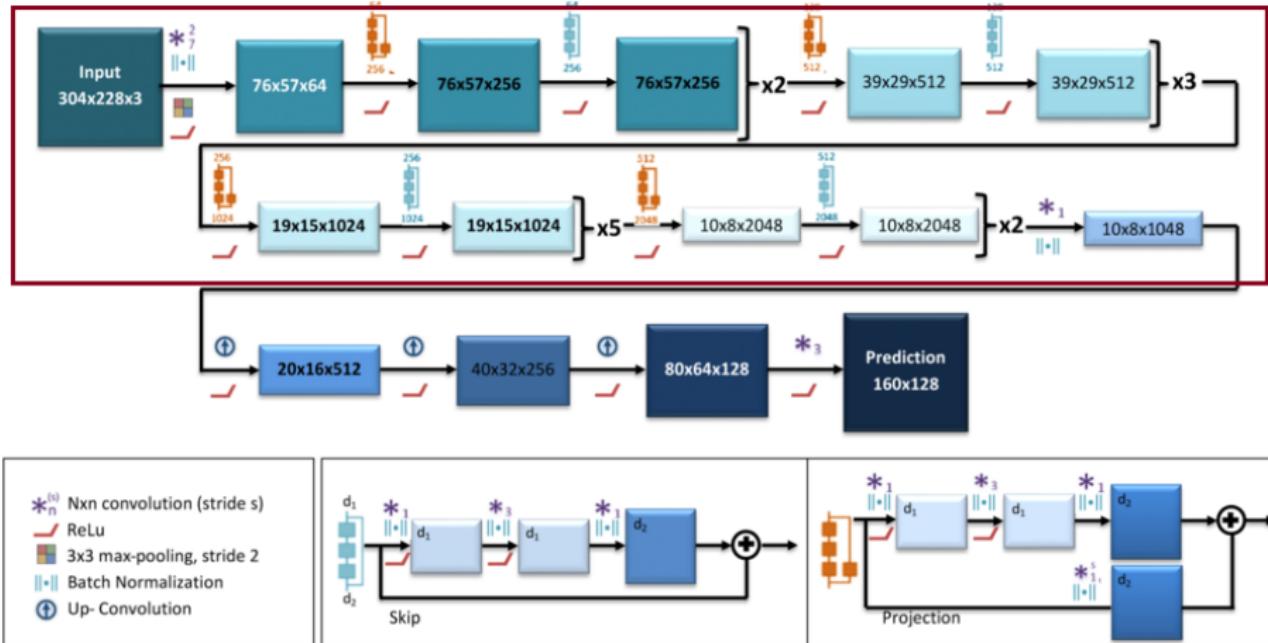


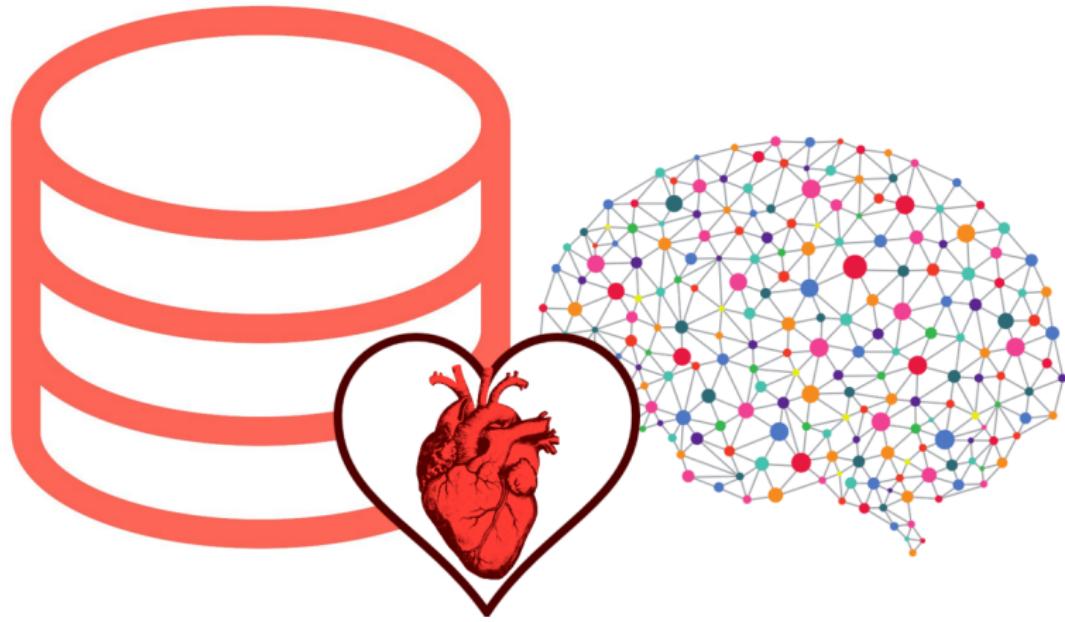
Depth from single monocular images

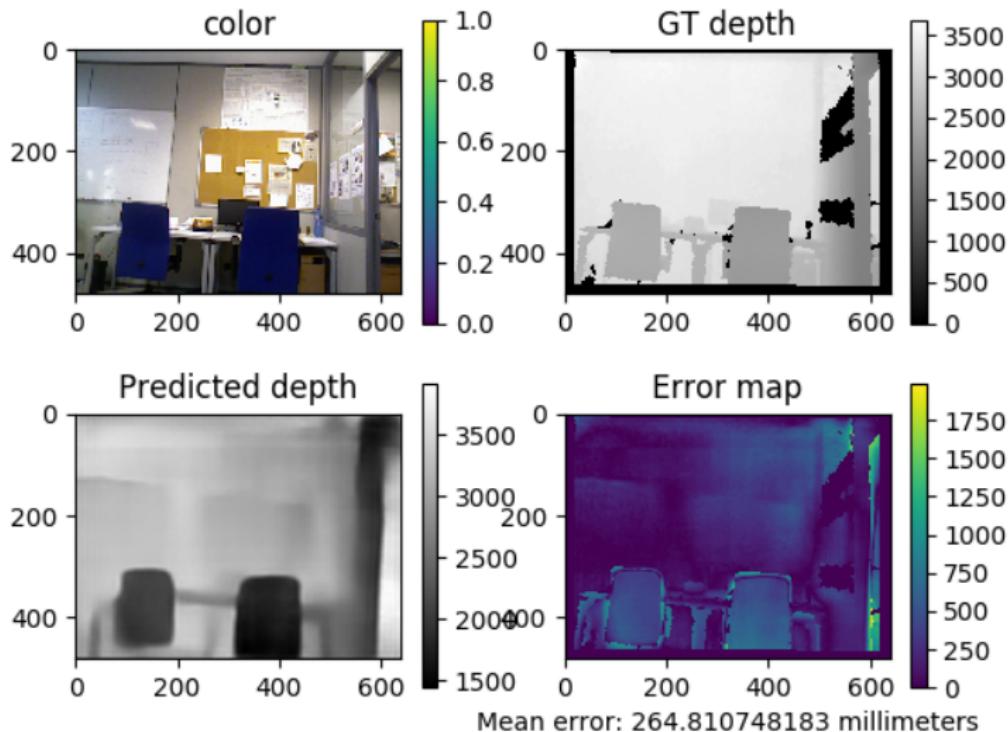


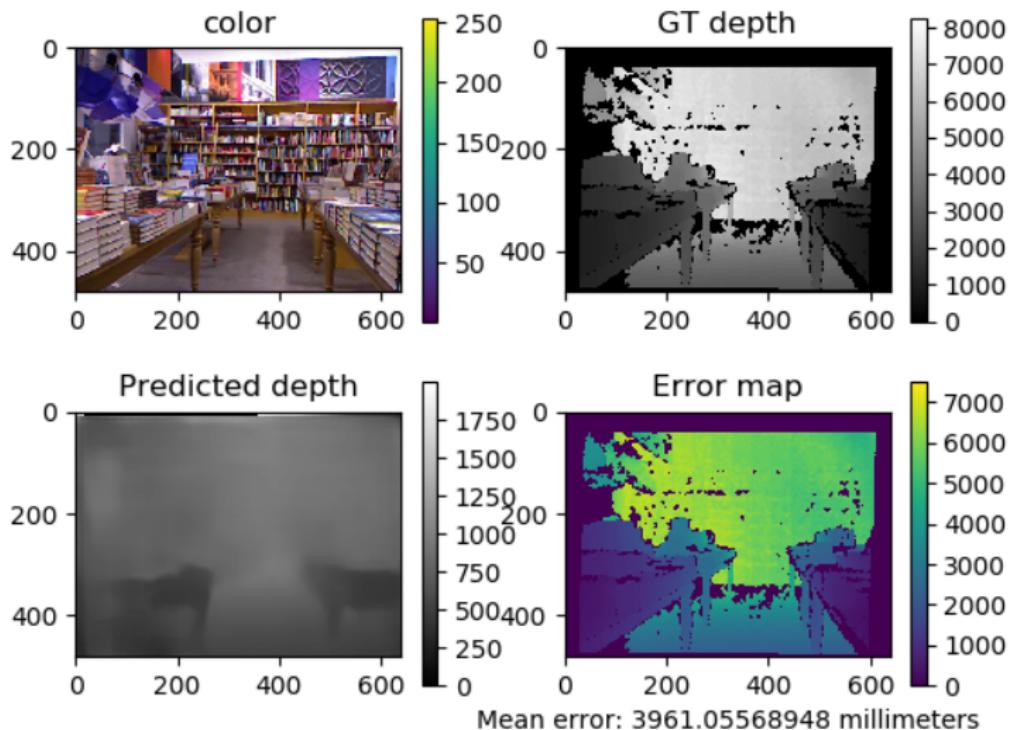
Universitat d'Alacant
Universidad de Alicante

Universidad de Alicante







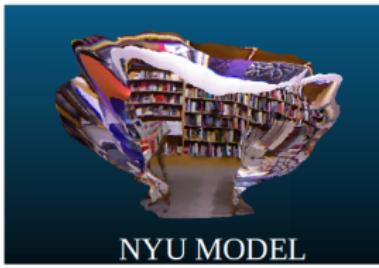


Results

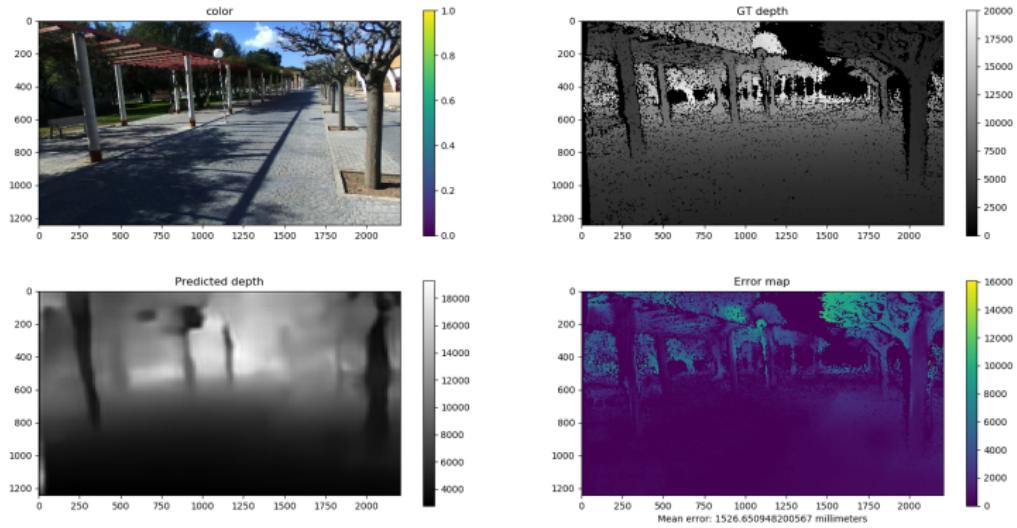


Universitat d'Alacant
Universidad de Alicante

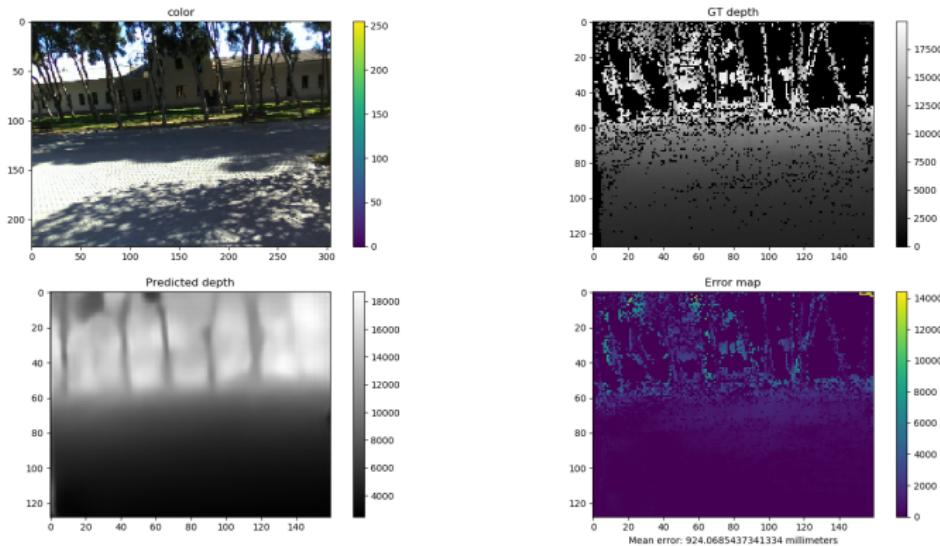
Universidad de Alicante



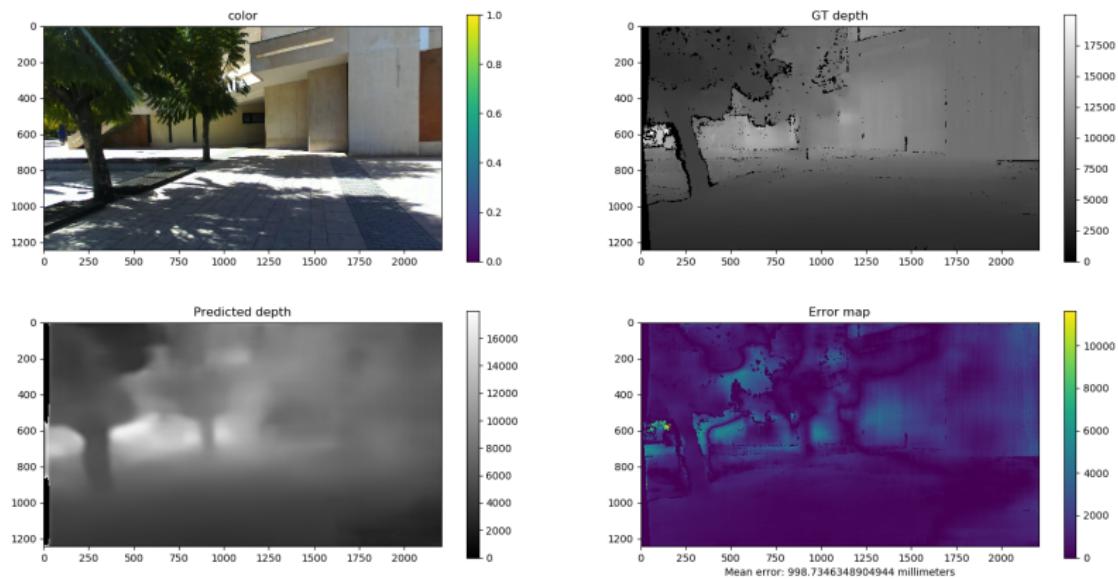
Results



Results



Results



Depth from single monocular images

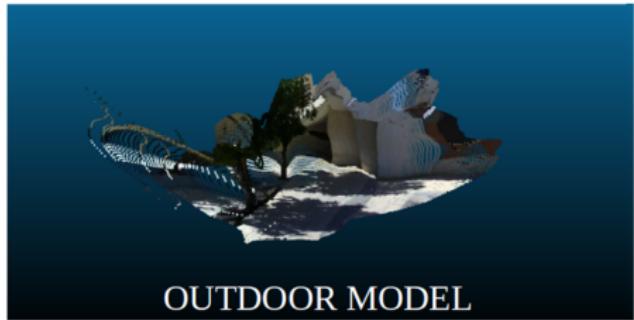


Universitat d'Alacant
Universidad de Alicante

Universidad de Alicante



GROUND TRUTH



OUTDOOR MODEL



Future Work

-  I. Armeni, A. Sax, A. R. Zamir, and S. Savarese.
Joint 2D-3D-Semantic Data for Indoor Scene Understanding.
ArXiv e-prints, Feb. 2017.
-  A. X. Chang, A. Dai, T. A. Funkhouser, M. Halber, M. Nießner,
M. Savva, S. Song, A. Zeng, and Y. Zhang.
Matterport3d: Learning from RGB-D data in indoor environments.
CoRR, abs/1709.06158, 2017.
-  B. Glocker, S. Izadi, J. Shotton, and A. Criminisi.
Real-time rgb-d camera relocalization.
IEEE, October 2013.

-  B.-S. Hua, Q.-H. Pham, D. T. Nguyen, M.-K. Tran, L.-F. Yu, and S.-K. Yeung.
Scenenn: A scene meshes dataset with annotations.
In *International Conference on 3D Vision (3DV)*, 2016.
-  A. Janoch.
The berkeley 3d object dataset.
Master's thesis, EECS Department, University of California, Berkeley, May 2012.
-  J. Martínez-Gómez, I. García-Varea, M. Cazorla, and V. Morell.
Vidrilo: The visual and depth robot indoor localization with objects information dataset.
The International Journal of Robotics Research, 34(14):1681–1687, 2015.

-  J. McCormac, A. Handa, S. Leutenegger, and A. J. Davison.
Scenenet rgb-d: 5m photorealistic images of synthetic indoor
trajectories with ground truth.
2016.
-  M. Menze and A. Geiger.
Object scene flow for autonomous vehicles.
In *Conference on Computer Vision and Pattern Recognition (CVPR)*,
2015.
-  P. K. Nathan Silberman, Derek Hoiem and R. Fergus.
Indoor segmentation and support inference from rgbd images.
In *ECCV*, 2012.

-  G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. Lopez.
The SYNTHIA Dataset: A large collection of synthetic images for semantic segmentation of urban scenes.
2016.
-  A. Saxena, M. Sun, and A. Y. Ng.
Make3d: Learning 3d scene structure from a single still image.
IEEE Trans. Pattern Anal. Mach. Intell., 31(5):824–840, May 2009.
-  T. Schöps, J. L. Schönberger, S. Galliani, T. Sattler, K. Schindler, M. Pollefeys, and A. Geiger.
A multi-view stereo benchmark with high-resolution images and multi-camera videos.
In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.



- S. Song, S. P. Lichtenberg, and J. Xiao.
Sun rgb-d: A rgb-d scene understanding benchmark suite.
In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

Depth from single monocular images



Universitat d'Alacant
Universidad de Alicante

Universidad de Alicante

nxm
in $n \times m$ convolution

2 x 2 un-pooling

ReLU

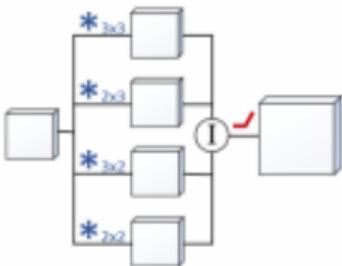
interleaving



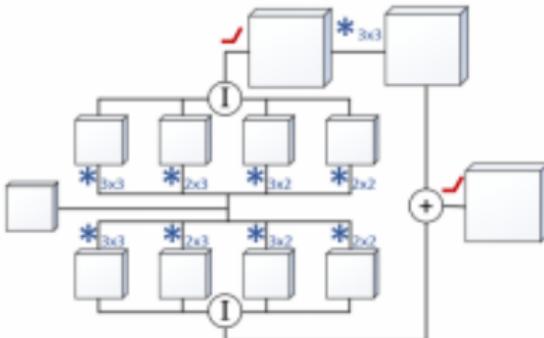
(a) up-convolution



(c) up-projection



(b) fast up-convolution



(d) fast up-projection

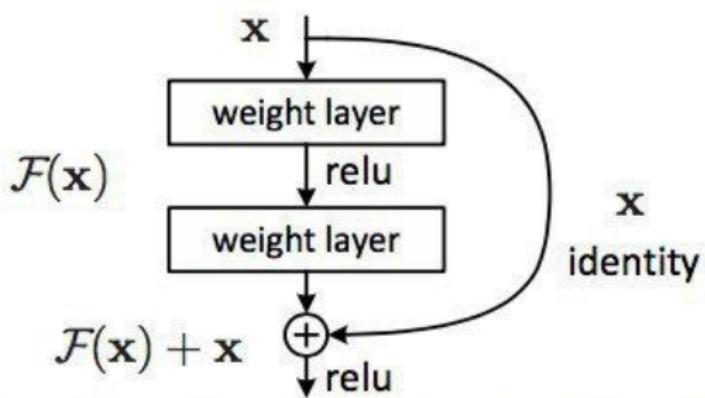


Figure 2. Residual learning: a building block.