



# Estimación de imágenes de profundidad a partir de imágenes a color usando aprendizaje profundo

Depth image estimation from color images using deep learning

Isaac De León Damas<sup>a</sup>, Boris Jesús Mederos Madrazo<sup>a\*</sup>, José Manuel Mejía Muñoz<sup>a</sup>

<sup>a</sup>Departamento de Ingeniería Eléctrica y Computación, Maestría en Ingeniería Eléctrica, Instituto de Ingeniería y Tecnología, México. \*Autor de correspondencia. Correo: boris.mederos@uacj.mx

---

## No. de resumen

2CP21-174

## Formato

Ponencia

## Evento

2.º Coloquio de Posgrados del IIT

## Presentador

Isaac De León Damas

## Tema

Procesamiento de Señales Digitales

## Estatus

Estudio en curso

## Fecha de la presentación

Noviembre 12, 2021

---

## Resumen

El objetivo de este proyecto es desarrollar un algoritmo eficiente capaz de llevar imágenes de color a su representación en profundidad, permitiendo reconstruir una representación tridimensional a partir de información RGB. Con RGB tomadas con una cámara monocular se pretende obtener una imagen de profundidad en la cual se utilizarán algunas herramientas, como son las redes neuronales convolucionales y optimización. El método entrena una red para la estimación de la profundidad monocular, utilizando una pérdida de reconstrucción de imagen. Esta red de estimación de profundidad está basada en la arquitectura general U-Net, es decir, una red de codificador-decodificador, con conexiones de salto, lo que permite representar tanto características abstractas profundas como información local. Como codificador se utilizó una red residual profunda ResNet18, que contiene 11 M de parámetros. Las capas internas del decodificador tienen funciones de activación de tipo ELU y la capa de salida tiene una sigmoide como función de activación. Para la función de costo de la red, se utilizaron dos términos: uno del error fotométrico y otro de suavidad que se combinaron para formar una función de costo global. En este trabajo se modificó la función de costo fotométrico de la red para usar otras métricas o funciones de costo que son robustas a desviaciones atípicas y grandes errores como la función de Barron.

**Palabras clave:** predicción de profundidad; cámara monocular; aprendizaje profundo; procesamiento de imágenes; red neuronal convolucional.



## Abstract

The objective of this project is to develop an efficient algorithm capable of carrying out color images to their representation in depth, allowing to reconstruct a three-dimensional representation from RGB information. From RGB images taken with a monocular camera, it is intended to obtain a depth image in which some tools such as convolutional neural networks and optimization will be used. The method trains a network for monocular depth estimation using an image reconstruction loss. This depth estimation network is based on the general U-Net architecture, that is, an encoder-decoder network, with skip connections, which allows representing both deep abstract characteristics and local information. A deep residual network ResNet18 was used as an encoder, containing 11M parameters. The inner layers of the decoder have ELU-type activation functions and the output layer has sigmoid as an activation function. For the network loss function, two terms will be used, a photometric error term and a smoothness term that will be combined to form a global loss function. In this work, the photometric loss function of the network was modified to use other metrics or loss functions that are robust to outliers and large errors such as the Barron function.

**Keywords:** depth prediction; monocular camera; deep learning; image processing; convolutional neural network.

## Entidad legal responsable del estudio

Universidad Autónoma de Ciudad Juárez.

## Financiamiento

CONACYT.

## Conflictos de interés

Los autores declaran que no existe conflicto de intereses.