

Depth Prediction and RGBD Images for Recognition

Yihui He, Metehan Ozten

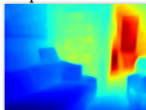
yihuihe@foxmail.com, m_ozten@umail.ucsb.edu

May 25, 2016

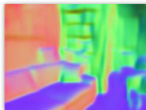
Related work and motivation



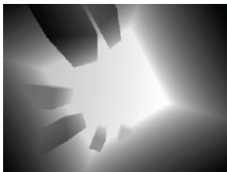
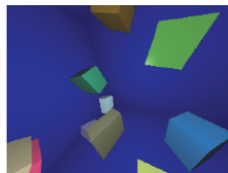
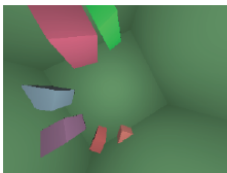
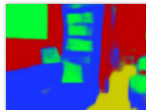
Depth



Normals



Labels



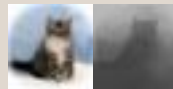
our project: depth estimation & Classification on RGBD images

implement previous work



Go further

(2) Build a RGBD CIFAR10 based on indoor depth knowledge



(3) Compare RGBD and RGB
 $label = f(RGBD)$
 $label = f(RGB)$

first part: implement previous work

infer depth from RGB image

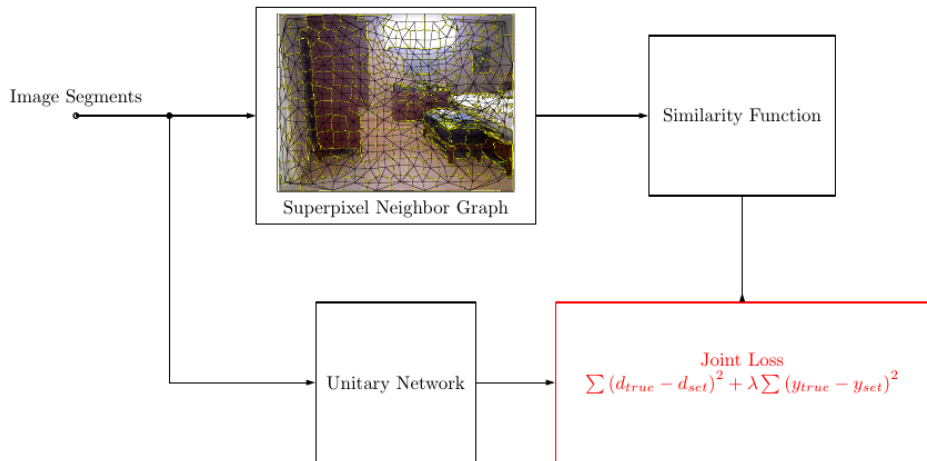
Infer depth from RGB image: Loss definition

At training time, we combine two objective function¹

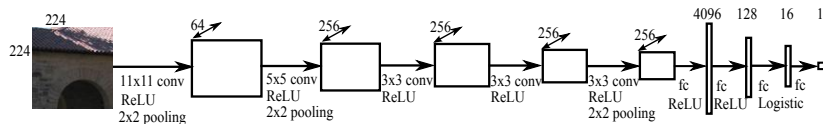
- 1 regress to ground truth depth image(Kinect, PrimeSense)
 $\sum_p (y_p - \hat{y}_p)^2$, p stands for pixel.
- 2 Similarity between superpixels. $R_{pq} = \sum_{k=1}^K \beta_k S_{pq}^{(k)}$
 β is trainable weight. S is similarity function.

¹Fayao Liu, Chunhua Shen, and Guosheng Lin. "Deep convolutional neural fields for depth estimation from a single image". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 5162–5170.

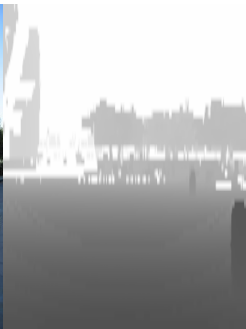
Architecture: Deep convolutional Neural Field



Infer depth from RGB image: Supervised part



using traditional CNN.



Compare performance with original paper

Method	Error (lower is better)			Accuracy (higher is better)		
	rel	log10	rms	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Our implementation	0.252	0.103	0.860	0.544	0.861	0.943
Original paper	0.230	0.095	0.824	0.614	0.883	0.971

Table: Sanity check (**Bold** is better)

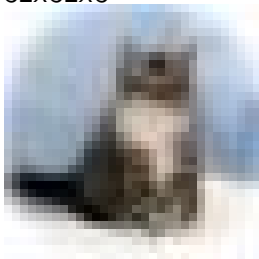
second part: go further

Classification on RGBD images

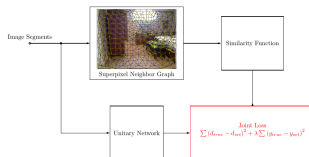
build RGBD CIFAR dataset



32x32x3



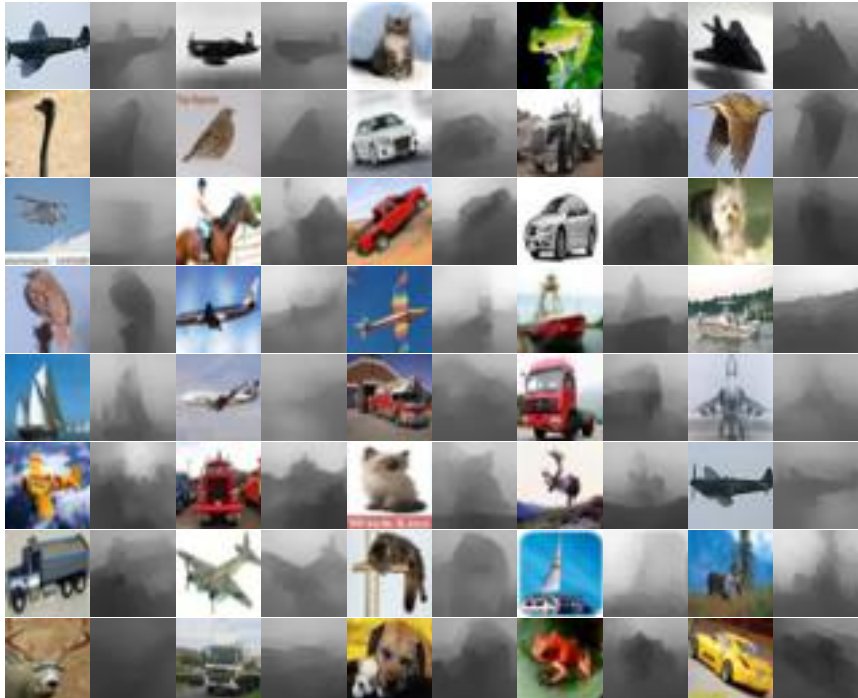
400x400x3



400x400x1



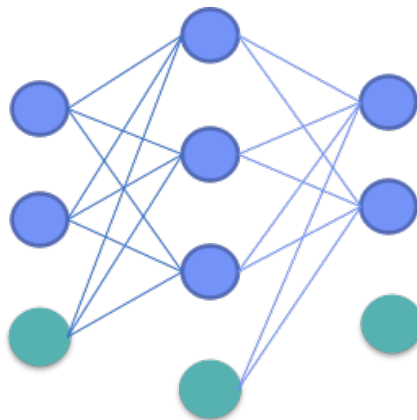
32x32x4



architecture

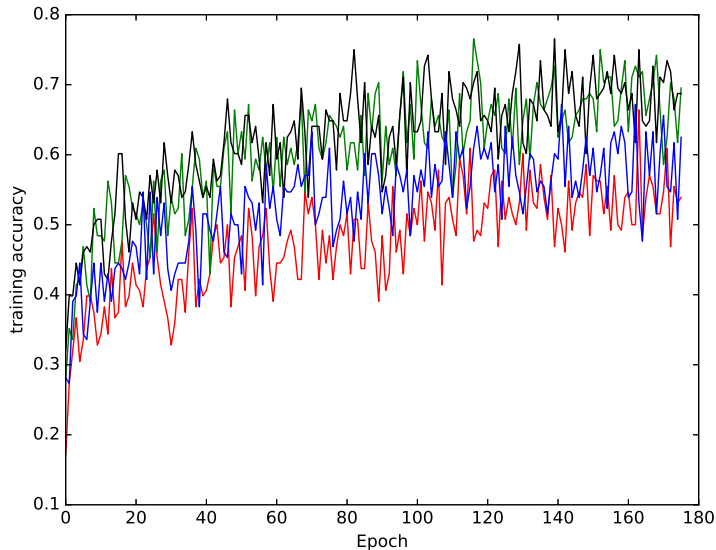


32x32x4

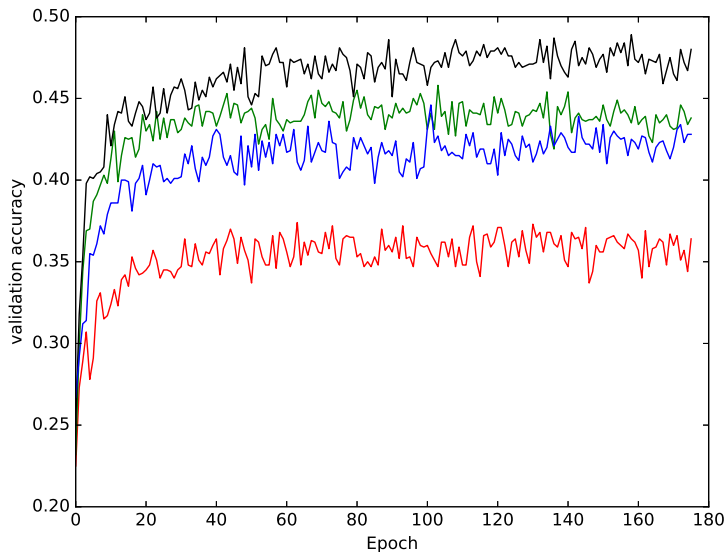


airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck

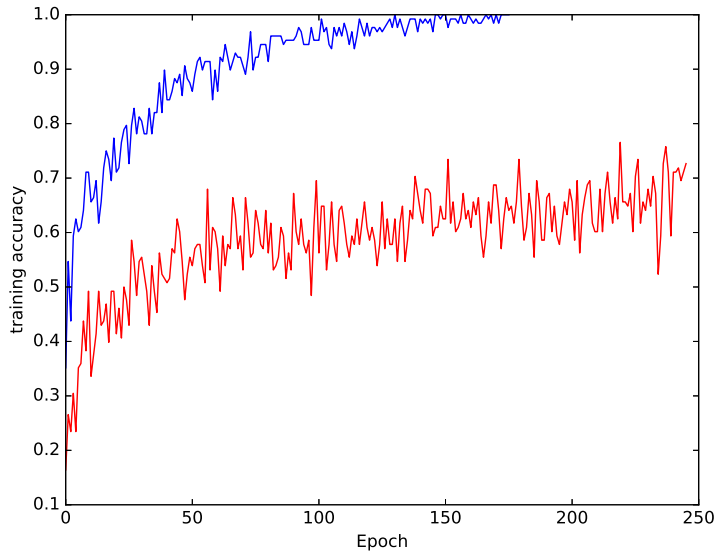
R vs G vs B vs D: training time



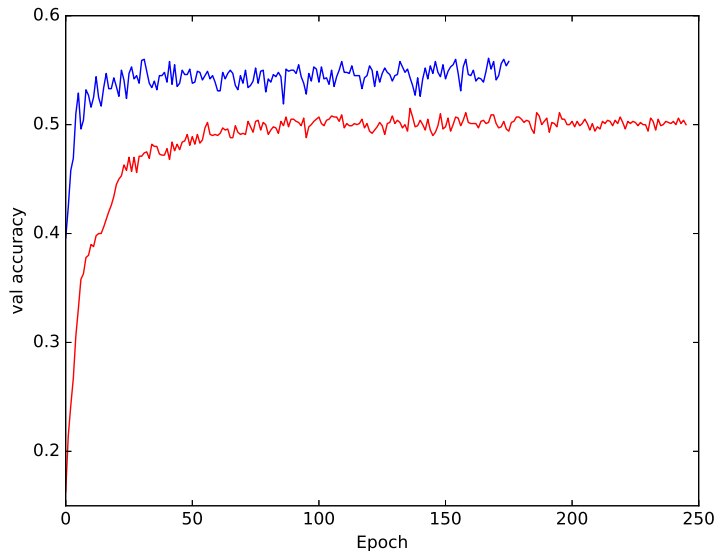
R vs G vs B vs D: testing time



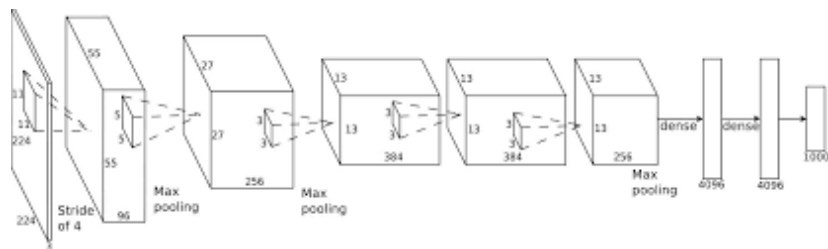
RGBD vs RGB: training time

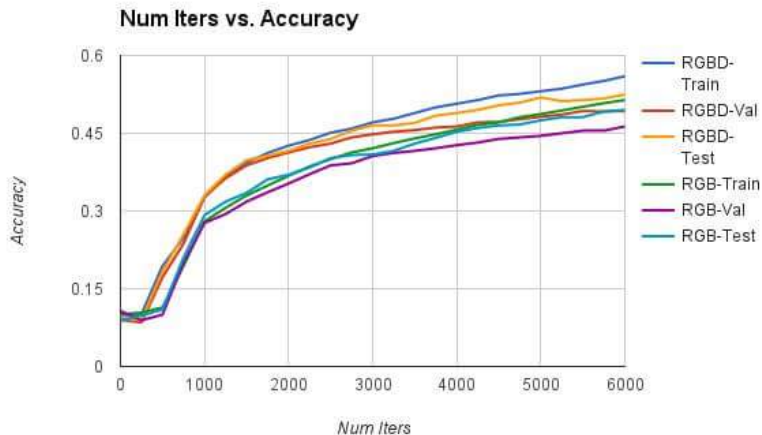


RGBD vs RGB: tges time



architecture





our contribution

- 1 reproduce previous work on depth estimation
- 2 create the first RGBD CIFAR10 dataset
- 3 define a new metric for depth prediction problem
- 4 prove that depth channel has a better feature representation
- 5 show that training on RGBD images can somehow improve accuracy

questions?²

²code, references, report and slides can be access here:
<https://github.com/yihui-he/Depth-estimation-with-neural-network>