















Crafting a multi-task CNN for viewpoint estimation Francisco Massa, Renaud Marlet, Mathieu Aubry Ecole des Ponts ParisTech



Method	Joint detector		Independent detector	
	mAP	mAVP24	mAP	mAVP24
Joint Regression 2D	49.2	15.7	51.6	16.4
Joint Regression 3D	49.6	17.1	51.6	17.4
Joint classification	48.6	21.1	51.6	20.5



Results

Dataset: Pascal 3D+ [4]

Summary of results and comparison to baselines

Method	mAVP24
DPM-VOC-VP [1]	13.6
Render for CNN [2]	19.8
Viewpoints & Keypoints [3]	31.1
Classif. Approach & AlexNet	19.3
+ our joint training	21.1
+ VGG16 instead of AlexNet	27.3
+ ImageNet data	34.4
+ synthetic data	36.1

Comparison with state of the art for different metrics

	[3]	Ours
mAVP4	49.1	55.4
mAVP8	44.5	51.3
mAVP16	36.0	40.6
mAVP24	31.1	36.1

New state of the art on Pascal 3D+ dataset [4] More results in the paper

Code available at imagine.enpc.fr/~suzano-f/bmvc2016-pose/

References

[1] Teaching 3D geometry to deformable part models. Pepik et al. CVPR 2012

[2] Render for CNN: Viewpoint estimation in images using CNNs trained with rendered 3D model views. Su et al. ICCV 2015

[3] Viewpoints and keypoints. Tulsiani and Malik. CVPR 2014 [4] Beyond Pascal: A benchmark for 3D object detection in the wild, Xiang et al, WACV 2014