University of Southampton

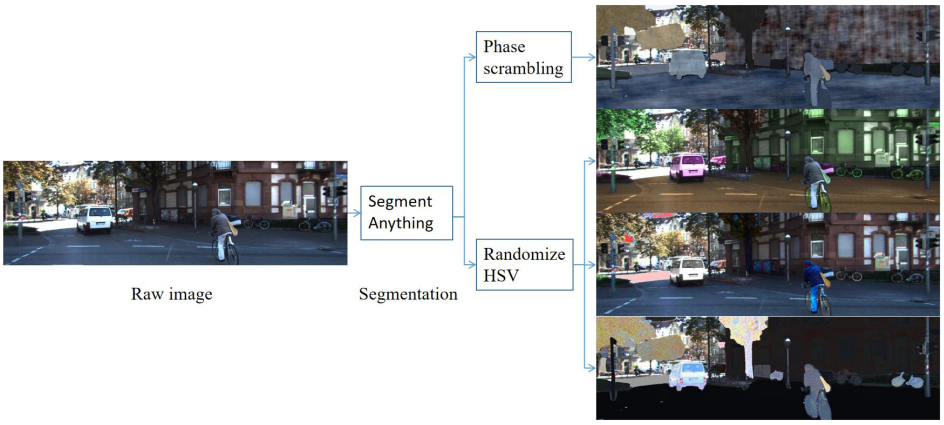
FEATURE CONTRIBUTION IN MONOCULAR DEPTH ESTIMATION

Hui Yu Lau, Srinandan Dasmahapatra, Hansung Kim
H.Y.Lau@soton.ac.uk

Introduction

- Bridging the gap between human and machine monocular depth perception (MDE)
- Understanding visual cues used by state-of-the-art MDE models
- Directed data-augmentation through understanding what models fail at
- Builds on previous work that measures existance of a single feature, we look at removal of feature
- Shape is assumed to be significant and basis of following features

Work Flow



Intervention workflow

Input images intervened upon to produce 4 intervention sets, first object segmentation to isolate objects, followed by separate interventions to remove image features: hue, saturation, intensity, texture.

Left: original image (A1: 0.9162) Right: intervened image (A1: 0.6039)

Value intervention on KITTI dataset

Compare drop in performance between 4 types of interventions. Contribution of a feature is measured by drop in performance following intervention

Experiments and results

- Texture is the most significant
- Intensity is the most significant colour feature, followed by saturation and hue

Left: Raw imput (0.9719)
Right: Intervened image (0.9706)

Left: Raw imput (0.9375)
Right: Intervened image (0.8862)

Hue

Saturation

Figure-a

Left: original (A1: 0.9897) Right: intervened (A1: 0.9651)

Figure-b

Left: original (A1: 0.9719) Right: intervened (A1: 0.9200)

Texture

Image type	A1	A2	A3	AbsRel	log10
KITTI original	0.9545	0.9896	0.9974	0.0796	0.0342
KITTI texture	0.8107	0.9388	0.9750	0.1468	0.0660
KITTI hue	0.9485	0.9901	0.9973	0.0833	0.0358
KITTI saturation	0.9449	0.9882	0.9967	0.0845	0.0363
KITTI intensity	0.8249	0.9526	0.9826	0.1332	0.0602
NYUv2 original	0.9714	0.9957	0.9990	0.0528	0.0229
NYUv2 texture	0.8751	0.9767	0.9946	0.1102	0.0477
NYUv2 hue	0.9703	0.9951	0.9988	0.0540	0.0234
NYUv2 saturation	0.9646	0.9937	0.9987	0.0588	0.0255
NYUv2 intensity	0.9265	0.9872	0.9966	0.0834	0.0362

Performance for various intervention sets

Conclusion

- Provided evidence that certain visual cues are used similarly between human and machine
- Discovered consistent roles of features in machine MDE

Future work

- More types of features that might extend beyond the image level
- An in depth study of performance distribution to further justify findings
- Causal discovery to identify dependencies between features