VirginiaTech Invent the Future



*equal contribution

- Scene Understanding Problem: Counting instances of object categories in everyday scenes
- Subitizing

- Visual Question Answering (VQA)







Counting Everyday Objects in Everyday Scenes

Prithvijit Chattopadhyay^{*,1}, Ramakrishna Vedantam^{*,2}, Ramprasaath R. Selvaraju², Dhruv Batra¹, Devi Parikh¹ ¹Georgia Institute of Technology, ²Virginia Tech

Datasets

Datasets

- PASCAL VOC 2007: 2501 train images, 2510 val images, 4952 test images and 20 object classes
- COCO: 82783 train images, 20252 val images (first half of COCO-val), 20252 test images (second half of COCO-val) and 80 object classes

Results

Metrics

- C_{ik} Ground truth count for class-k and image-i
- $\hat{c_{ik}}$ Predicted count for class-k and image-i

$$RMSE_k = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\hat{c_{ik}} - c_i)}$$

$$2 \quad relRMSE_k =$$

$$\boxed{\frac{1}{N} \sum_{i=1}^{N} \frac{(\hat{c_{ik}} - c_{ik})}{c_{ik} + 1}}$$

$$\sum_{i=1}^{k} \frac{(c_{ik} - c_{ik})^2}{c_{ik} + 1}$$

Root-mean Squared Error **Relative Root-mean** Squared Error

Models	mRMSE	mRMSE-nz	mrelRMSE	mrelRMSE-nz
Detection (Baseline)	0.49(0.00)	2.78(0.03)	0.20(0.00)	1.13(0.01)
Glancing (Baseline)	0.42(0.00)	2.25(0.02)	0.23(0.00)	0.91(0.00)
Aso-sub (Baseline)	0.38(0.00)	2.08(0.02)	0.24(0.00)	0.87(0.01)
Seq-sub (Proposed)	0.35(0.00)	1.96(0.02)	0.18(0.00)	0.82(0.01)

Counting performance on COCO Count-test split. nz = non-zero counts



Bottle GT: 8 Detect: Glance: Aso-sub: 10 Seq-sub:



Elephant GT: 16 Detect: 3 Glance: 9 Aso-sub: 22 Seq-sub: 17

Qualitative Examples

IEEE 2017 Conference on **Computer Vision and Pattern** Recognition



Applications

Visual Question Answering

- 10.28% questions in VQA are counting-Q
- 7.07% questions in COCO-QA are counting-Q
- Count-QA: Subset of counting questions in VQA + COCO-QA



(Fukui et al. 2016)

Seq-sub (Proposed)

Improving Object Detection

• Detectors are typically operated at some threshold which is usually set on a global basis

1.81(0.09)

• Use counting to set per-image thresholds, based on count estimate

Method	mF(%)	
Category-wise Threshold	15.26	
Ground Truth (oracle)	20.17	
Seq-sub (Proposed)	17.00	

Evaluation Metric: mean F-measure (mF)