CONTEXT-AWARE AUTOMATIC OCCLUSION REMOVAL

Kumara Kahatapitiya, Dumindu Tissera, Ranga Rodrigo

k.kahatapitiy@cs.stonybrook.edu, dumindutissera@gmail.com, ranga@uom.lk

University of Moratuwa, Sri Lanka

MOTIVATION

- Existing image enhancement techniques for occlusion removal:
  - Domain-specific. Eg: shadow removal, image de-raining
  - Require manual-annotation.
- No work tries to capture occlusions based on image-context.

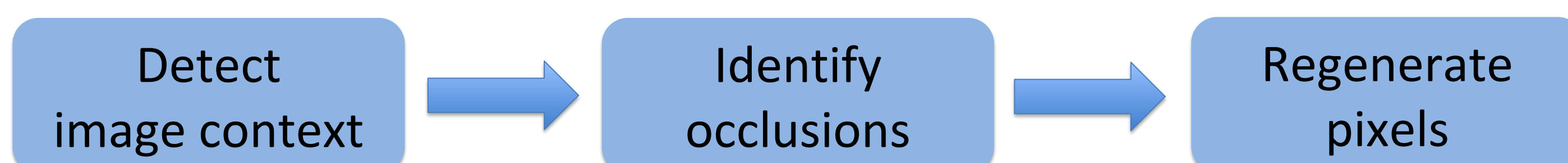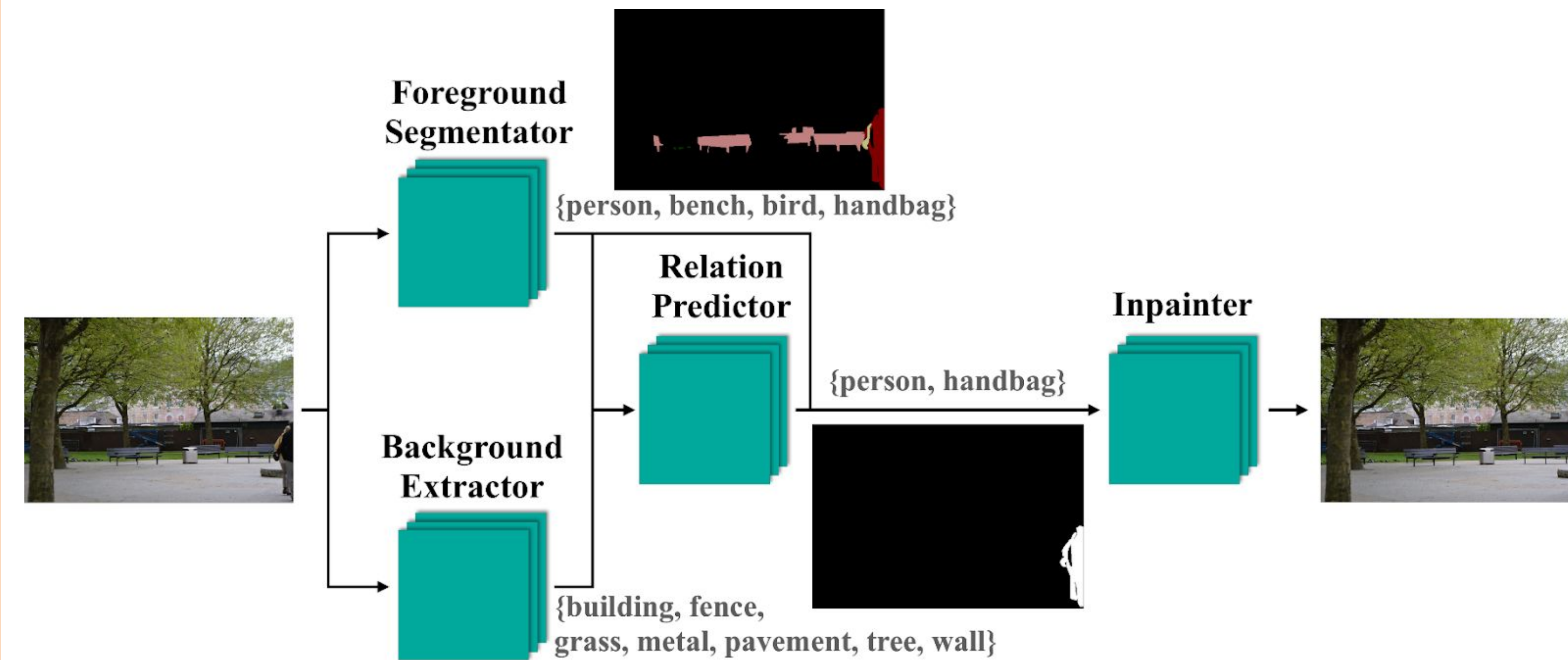[Qian *et al.*]   [Wang *et al.*]

CHALLENGES

- How to capture image context in a generic domain?
  - Highly varying and complex.
  - Subjective in human perception, required to be captured objectively.
- How to evaluate?
  - No baseline.
  - No dataset annotating image-context and respective occlusions.

PROBLEM FORMULATION

- Making intelligent decisions:
  - Identifying image context based on background and foreground objects.
  - Detecting objects not related to image context as occlusions.
- Producing a visually-pleasing output:
 - Replacing the pixels related to occlusions coherently.

Detect image context → Identify occlusions → Regenerate pixels

SYSTEM ARCHITECTURE

Foreground Segmentator

{person, bench, bird, handbag}

Relation Predictor

Inpainter

{person, handbag}

Background Extractor

{building, fence, grass, metal, pavement, tree, wall}

IMPLEMENTATION

- Data used for training sub-networks:
  - Foreground Segmentator - foreground segmentations.
  - Background Extractor - background class labels.
  - Relation Predictor - image captions.
  - Inpainter - images and random masks.
- Original corpus vs modified corpus of image captions for Relation Predictor.
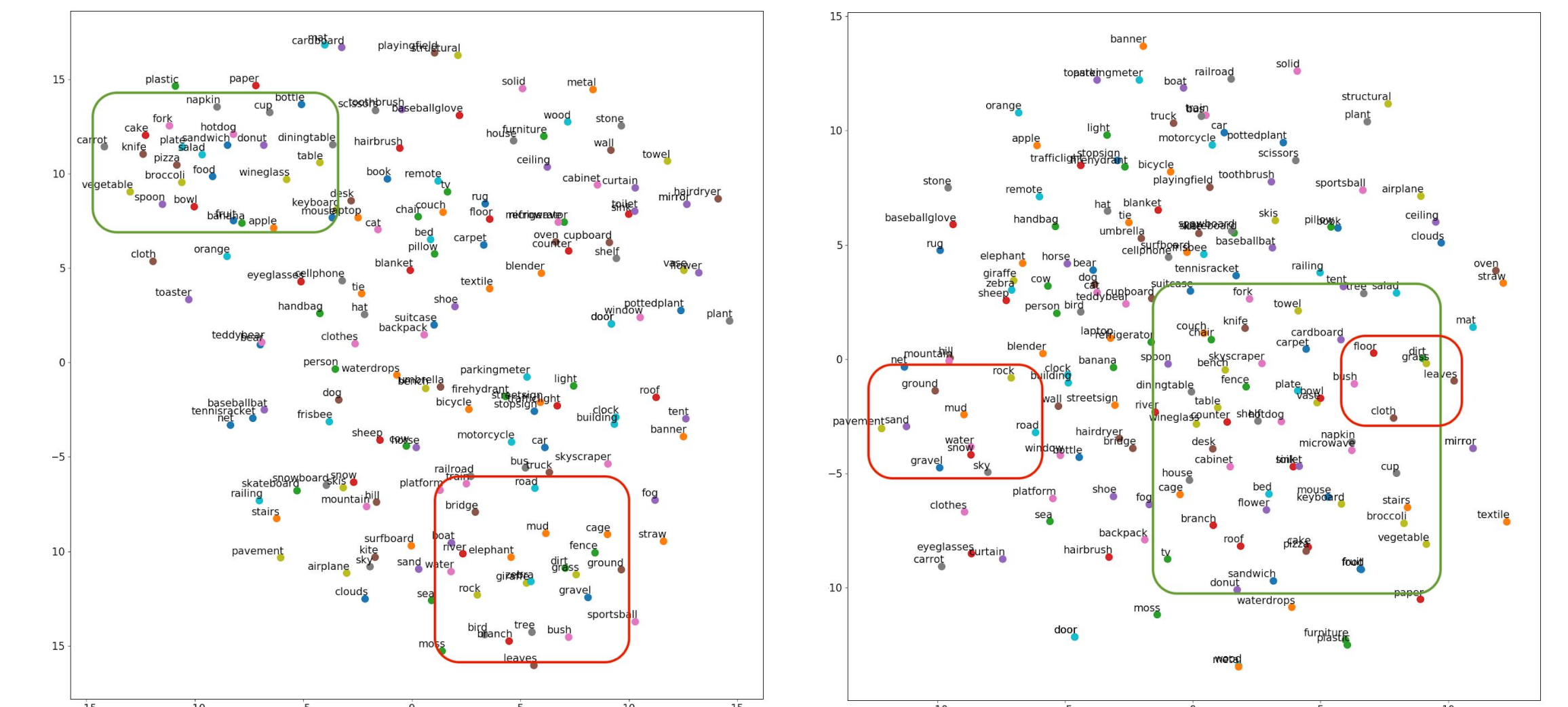- Random masks for Inpainter.
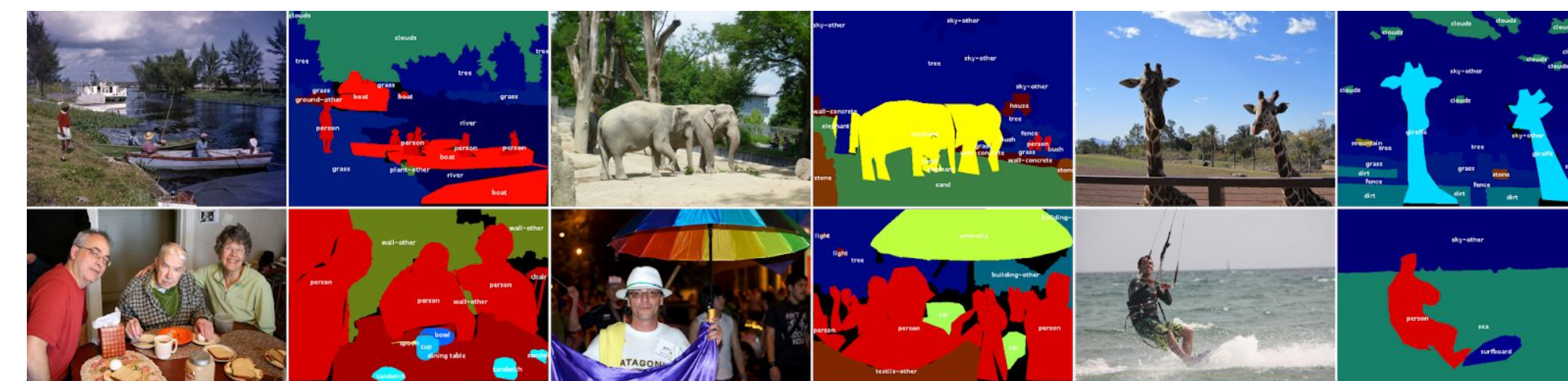
COCO-Stuff

- Combining sub-networks aggregate errors.
- We evaluate:
  - What is removed?
  - How good is the reconstructed image?
- Direct evaluation of end-to-end system is impossible:
  - Dataset has no annotations on image context or occlusions.
  - Dataset has no image-pairs with and without occlusions.

RESULTS

- Effectiveness of Word-Embeddings

- User Study

Visually-pleasing	
Positive	992/1245 **79.7%**
Negative	253/1245 **20.3%**

Relation	
Precision	39.03%
Recall	17.46%

- Visually-Pleasing nature

CONCLUSION

- We establish a baseline for context-aware automatic occlusion removal in a generic domain, even with the lack of a relation based dataset.
- Although our approach learns meaningful relationships between object classes and utilizes hand designed algorithms to decide on occlusions, how humans perceive it can be different
- As future work, we hope to develop a dataset that captures human annotations on object relations, which will enable end-to-end training of such networks.