Supplementary Material For Learning Attention Propagation for Compositional Zero-Shot Learning

Muhammad Gul Zain Ali Khan[1,3,4] Muhammad Ferjad Naeem[2]   Luc Van Gool[2]
A. Pagani[1,3]   Didier Stricker[1,3,4]   Muhammad Zeshan Afzal[1,3,4]
[1]DFKI,   [2]ETH Zürich,   [3]TU Kaiserslautern,   [4]MindGarage

We share some interesting images of compositions shown in Table. 2 of our paper in Figure. 1 and Figure. 2. We show ground truth composition of each image at the top of image in dark blue color. The first column from the left contains examples of images from the query compositions in Table. 2 of the our paper. The second column from the left contains images from some of the Top 5 compositions in Table. 2 of the our paper. We also label shared **objects** of query and value images. Shared objects are bounded by dark green boxes. Yellow text on these bounding boxes represent the label of shared **object**.

A shared object is an object that can be seen and is shared across the query images and value images. Figure. 1 shows results for CGQA [2] dataset and Figure. 2 shows results for MIT-States dataset. The majority of our shared images do not contain compositions that share any primitives yet contain shared objects. This is done to demonstrate that there exists complex interdependency structure between compositions in Table 2 of the paper that goes beyond shared primitives. Our approach finds optimal propagation routes between compositions to propagate knowledge between them. Figure. 1 shows "Red Floor" and "Carpeted Floor" images have "Carpet" in them. In this example, "Carpet" is a shared object. Compositions "Cluttered Kitchen", "Cluttered Counter" and their query composition "Blue Mug" have object "Mug" in them. Compositions "Red Chair", "Stripped Chair" and their query composition "Yellow Wall" have a "Chair" in them. All of these compositions share objects that have visual similarities.

Figure. 2 shows compositions "Pureed Fish" and its query composition "Sliced Salmon" have sliced fish. Composition "Broken Log", "Peeled Log" and their query composition "Weathered Redwood" have logs in them. Compositions "Old Log", "Old Boat" and their query composition "Rusty Bridge" have some form of wood or water in them. Composition "Old Boat" is a boat made out of wood and "Rusty Bridge" either has logs or is made of wood. All of these compositions share objects that have visual similarities.

We proposed an approach that can discover complex inter-dependency structure between compositions. Our approach finds critical propagation routes between the compositions to propagate knowledge between them. Results in Figure. 1 and Figure. 2 shows that our approach can find complex interdependency structures between compositions that do not share primitives yet share some aspects. Our approach can find these shared aspects of compositions by exploiting self-attention on word embeddings of compositions.

CGQA

Query Image

Value Images

Red Floor — Carpet

Red Floor — Carpet

Carpeted Floor — Carpet / Carpet

Carpeted Floor — Carpet

Blue Mug — Mug

Blue Mug — Mug

Cluttered Kitchen — Mugs

Cluttered Counter — Mug

Yellow Wall — Chair

Yellow Wall — Chair

Red Chair — Chair
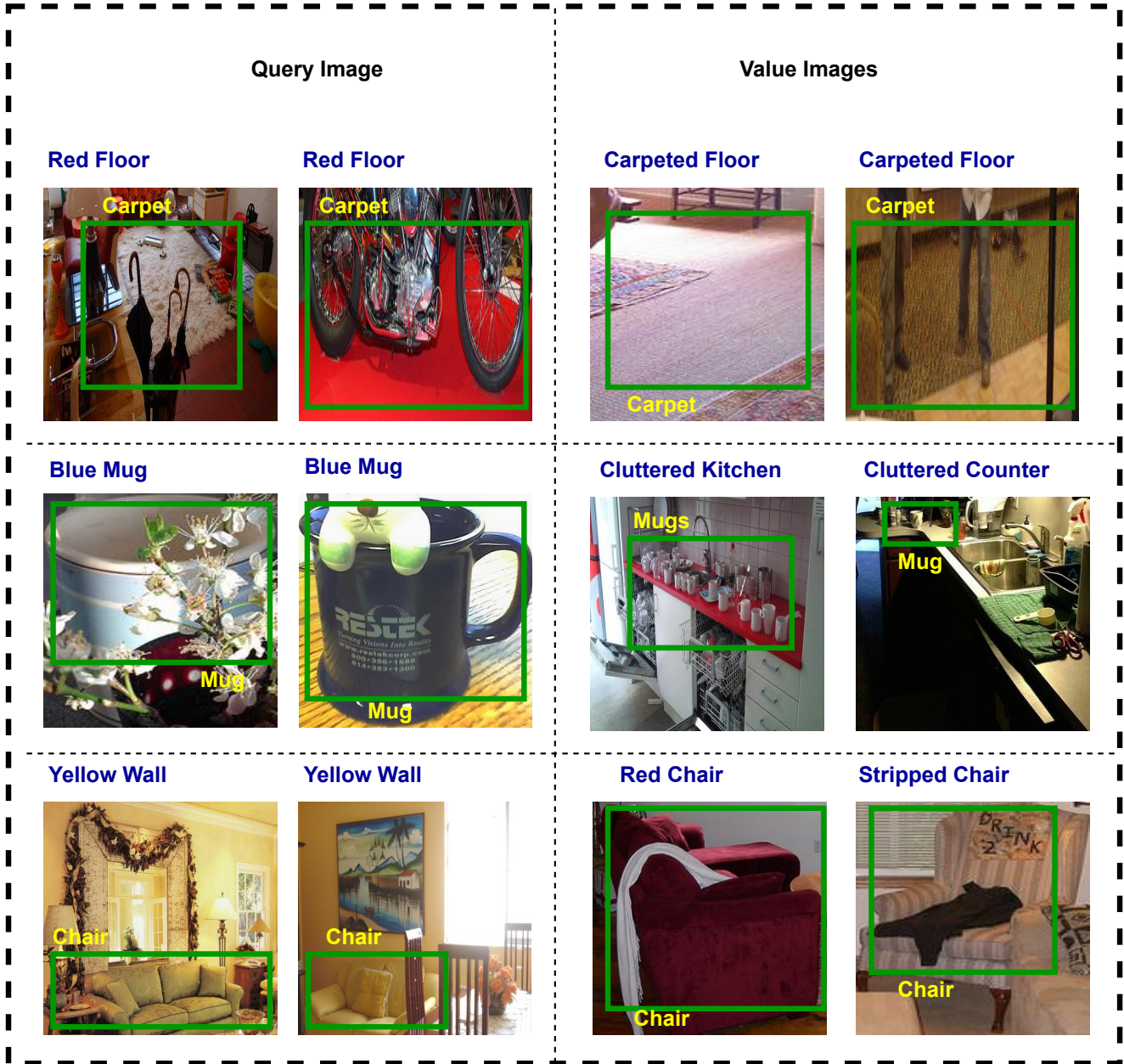
Stripped Chair — Chair

Figure 1. We share qualitative images of Queries and some of their Top 5 compositions from Table. 2 of our paper. This figure contains results of compositions from CGQA [2] dataset. We show ground truth composition of each image at the top of image in dark blue color. First column from left contains example of images from the query compositions of Table. 2 of our paper. Second column from left contains images from some of Top 5 compositions of Table. 2 of our paper. We also label shared **objects** of query and value images. Shared objects are bounded by dark green boxes. Yellow text on these bounding boxes represent the label of shared **object**.

MIT-States

Query Image

Value Images

Sliced Salmon

Sliced Salmon

Pureed Fish

Pureed Fish

Sliced Fish

Sliced Fish

Sliced Fish

Sliced Fish

Weathered Redwood

Weathered Redwood

Broken Log

Peeled Log

Logs

Logs

Log

Logs

Rusty Bridge

Rusty Bridge

Old Log

Old Boat

Wood

Wood

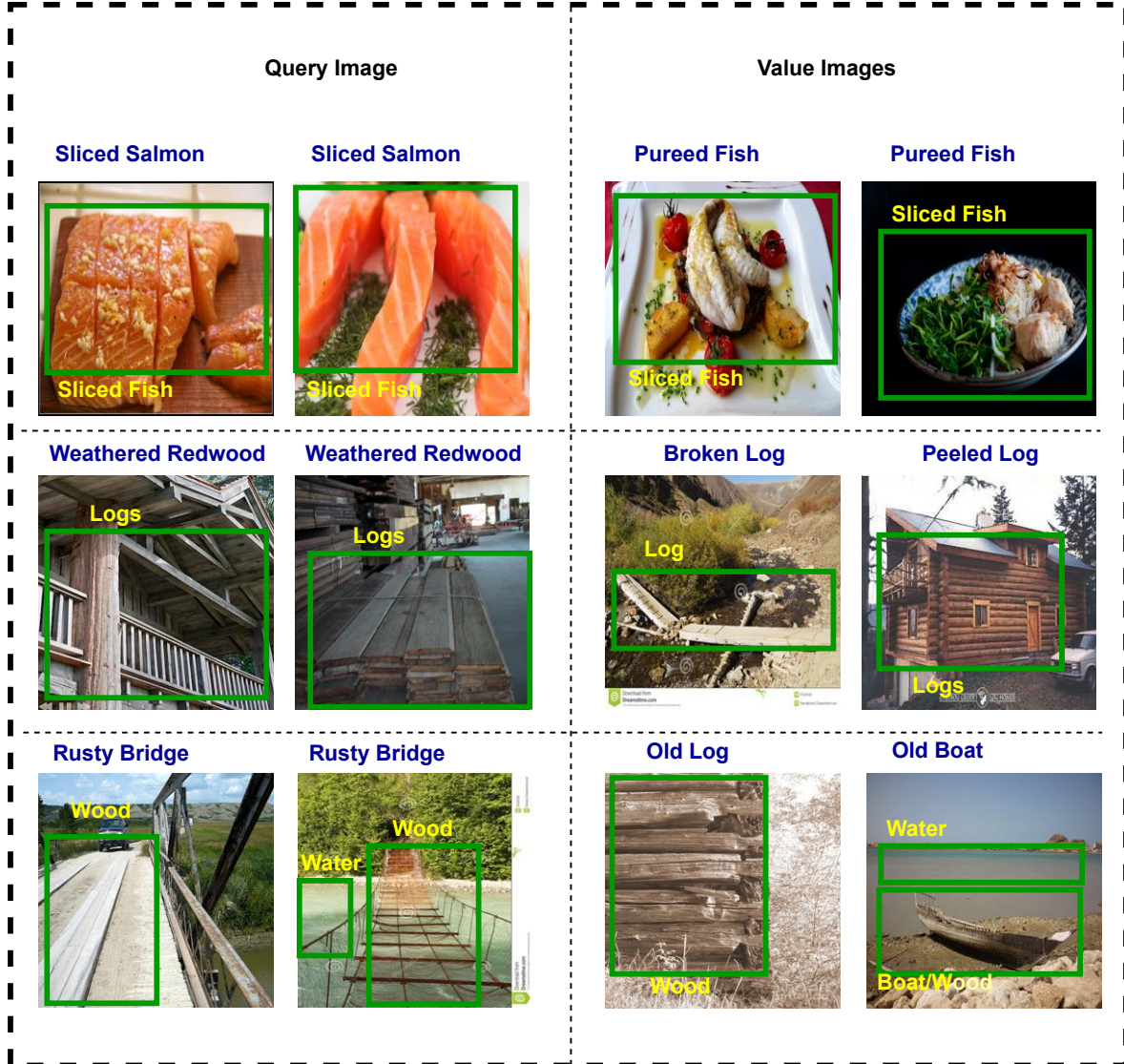Water

Water

Water

Wood

Boat/Wood

Figure 2. We share qualitative images of Queries and some of their Top 5 compositions from Table. 2 of our paper. This figure contains results of compositions from MIT-States [1] dataset. We show ground truth composition of each image at the top of image in dark blue color. First column from left contains example of images from the query compositions of Table. 2 of our paper. Second column from left contains images from some of Top 5 compositions of Table. 2 of our paper. We also label shared **objects** of query and value images. Shared objects are bounded by dark green boxes. Yellow text on these bounding boxes represent the label of shared **object**.

References

[1] Phillip Isola, Joseph J Lim, and Edward H Adelson. Discovering states and transformations in image collections. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1383–1391, 2015.

[2] Muhammad Ferjad Naeem, Yongqin Xian, Federico Tombari, and Zeynep Akata. Learning graph embeddings for compositional zero-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 953–962, 2021.