GROUND TRUTHS

- **2014**

  We therefore first collect a dataset of 100 light fields using the Lytro light field camera. The dataset consists of 60 indoor scenes and 40 outdoor scenes. For each data, we ask three individuals to manually segment the saliency regions from the all-focus image. The results are deemed ground truth only when all three results are consistent (i.e., they have an overlap of over 90%)

- **2015 - WSC**

  same GT as 2014 (same dataset)

- **2015**

  same GT as 2014 (same dataset)

- **2017**

  We create ground truths by human subjects. Five subjects, 3 males and 2 females, are asked to draw bounding contours around the objects/regions that attract them in the image. It is noticed that some regions marked by the subjects have inconsistencies in terms of being salient ground truths such as the regions where there are multiple objects in the scene. Hence, we set the pixel value to 1 if at least 3 subjects agree that it belongs to a salient region and zero otherwise. Since one should respond uniformly within the whole region [28], finally, one participant uses Adobe Photoshop to segment the salient region maps manually from each image

- **2018**

  manually labeled ground truths

- **2019 -AFNet**
- **2019 - single view**

- **2019 - DL4LFSD**

  To obtain the pixel-wise ground truth, we manually label the images using a custom segmentation tool. We first draw the coarse boundary along the salient objects and then check the segmentation results to further refine the boundaries until we obtain final accurate annotation.

- **2020**

  The micro-lens image array is firstly used as input for light field saliency detection. Then, we annotate per-pixel ground truth for each central viewing image.

- **2022**

  DUTLF-V2 [22], which includes rich data types (RGB images, corresponding manually labeled ground truths, depth maps, stacks of focal slices, and multi-view images).

- **2023**

  We employ thirty participants divided into ten groups to conduct our subjective experiments. Every three participants within a group jointly identify salient objects in light fields, which are then cross-validated by other groups. Only the object confirmed by over 80 percent of participants will be deemed the positive label to ensure a high consistency. The annotation quality is an essential factor influencing the lifespan of datasets. To this end, we resort to professional annotators to obtain diverse types of annotations (see Fig. 3), e.g., scribble annotations, bounding boxes, object-/instance-level annotations, and edge annotations. The rich annotations in our dataset are expected to facilitate many potential tasks, not limited to SOD. Besides, since our annotations are labeled on high-resolution (2022×1404) images, as shown in

Fig. 5, every detail of salient objects can be traced, which benefits the performance improvement of existing SOD algorithms