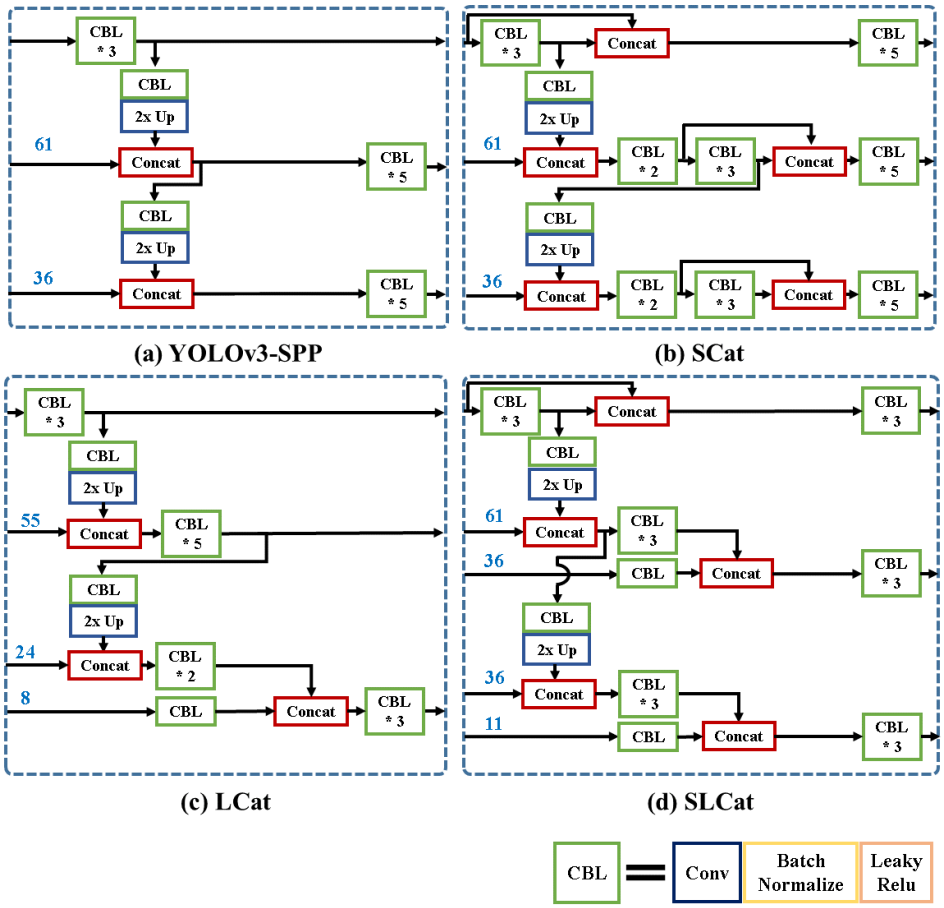

Figure 1: Details of baseline and our proposed models: (a)YOLOv3 with SPP module (b) SCat(Short-skip Concatenation), (c)LCat(Long-skip Concatenation), and (d) SLCat(Short-Long-skip Concatenation)

A Model Architecture

Fig. 1 shows the details about our proposed model. In Short-skip Concatenation(SCat) model, we add one fusion module on first scale prediction layer, and split five convolutional layers in YOLOv3-SPP into two parts at the second and third prediction layer, then use the skip-connection layer on the neck part of the model. Each concatenated skip-connection is followed by five additional convolution operations to process the features more useful to the detection.

Second model is LCat which uses longer skip layer than the baseline model. While YOLOv3-SPP brings the local feature maps from layer 61 and 36, we draws lower-level features with one more concatenation layer from layer 8, as shown in Fig. 1c.

Last model is SLCat(Fig. 1d) which combines previous two approaches, fusing the information from the backbone and neck part. Using extra three convolutional layer after concatenation module, we prevent the unnecessary features from decreasing the performance. Unlike LCat model, we follow the same concatenation from backbone as YOLOv3-SPP, and use features from layer 36 at 2nd layer as well. Then we add one fusion module from backbone layer 11.

B Sample Image

Real Fisheye Image Fig. 2 shows sample images captured from FE185C057HA-1 fisheye lens camera with 185° field of veiw .

Fisheye and Projected Images Fig. 3 appears fisheye and projected images with detection results from Fisheye-CityScape(top), Fisheye-KITTI(middle), and Fisheye-Dongseongno(bottom).

C Implementation Details

Synthetic Fisheye-KITTI We evaluate our methods on KITTI 2012 object detection dataset with 3.8K training images at input size 512.

Each model is trained using Adam optimizer with momentum 0.937 and weight decay 5e-4. Initial learning rate is 1.9e-4 and we adopt cosine decay learning rate scheduling strategy.

During training, each network is trained for 350 epochs with total batch size 24 on one GPU.

Synthetic Fisheye-CityScape We evaluate our methods on CityScape by generating bounding boxes with 3.8K training images at input size 512.

Each model is trained using Adam optimizer with momentum 0.843 and weight decay 3.6e-4. Initial learning rate is 1.9e-4 and we adopt cosine decay learning rate scheduling strategy.

During training, each network is trained for 350 epochs with total batch size 24 on one GPU.

Real Fisheye-Dongseongno We evaluate the proposed methods on our own dataset from fisheye lens with 2500 training and 600 test images at input size 640.

Each model is trained using Adam optimizer with momentum 0.843 and weight decay 3.6e-4. Initial learning rate is 1.9e-4 and we adopt cosine decay learning rate scheduling strategy.

During training, each network is trained for 350 epochs with total batch size 20 on one GPU.

Figure 2: Sample images from 185° FOV FE185C057HA-1 fisheye lens with a 2/3 inch sensor

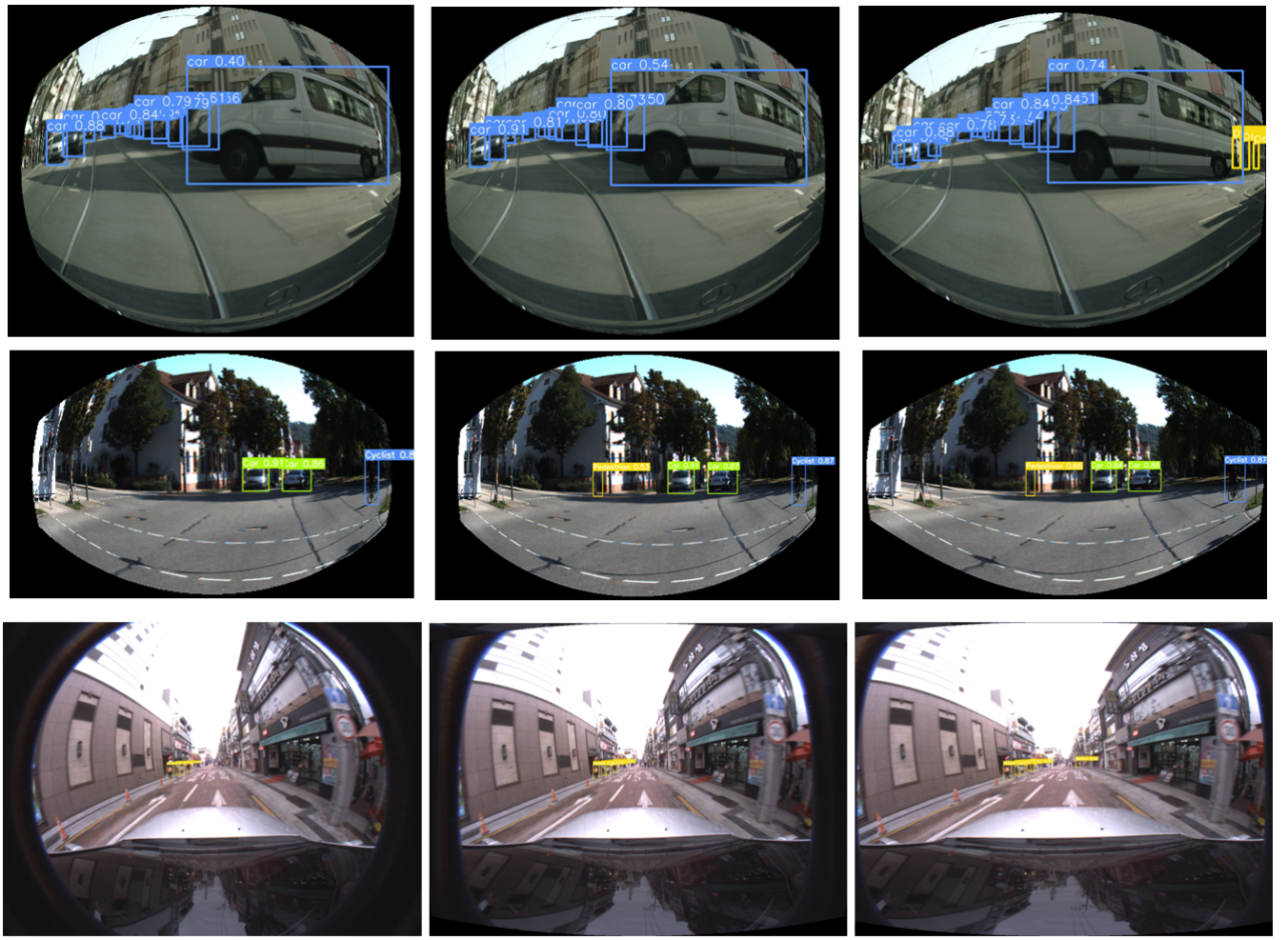

Figure 3: Sample images of fisheye, basic-spherical and expandable spherical projection with detection results from baseline model. Sample image from Fisheye-CityScape(top), Fisheye-KITTI(middle), Fisheye-Dongseongno(bottom)