Supplementary Material: RadarDistill: Boosting Radar-based Object Detection Performance via Knowledge Distillation from LiDAR Features

Geonho Bang[1*] Kwangjin Choi[1*] Jisong Kim[1] Dongsuk Kum[2] Jun Won Choi[3†]
[1]Hanyang University, Korea [2]KAIST, Korea [3]Seoul National University, Korea

[1]{ghbang, kjchoi, jskim}@spa.hanyang.ac.kr [2]dskum@kaist.ac.kr [3]junwchoi@snu.ac.kr

This supplementary material contains additional architecture details and experimental results for the proposed RadarDistill.

1. Detailed Architecture

Fig. 1 depicts the detailed structure of CMA. The *Down block* extracts the features through a 3×3 deformable convolution [1] followed by two ConvNeXt v2 [2] blocks. Each ConvNeXt v2 block consists of a 7×7 depth-wise convolution followed by layer normalization, a 1×1 convolution layer, GELU, and another 1×1 convolution layer. The first 1×1 convolution layer expands the feature channel from 256 to 1024 while the second 1×1 convolution layer reduces it back to 256. The Up block includes a 4×4 transposed convolution with a stride of 2 for dilation, followed by batch normalization and GELU. The Aggregation module combines two input features through concatenation followed by a 1×1 convolution layer, batch normalization, and GELU.

Fig. 2 presents the detailed structure of DenseEnc. The Conv2D-BN-ReLU block consists of a 3×3 convolution layer with a stride of 2, followed by Batch Normalization (BN) and Rectified Linear Unit (ReLU). This block reduces the resolution of F^l_{mod} to 1/16. The 2D Conv Block consists of six layers of Conv2D-BN-ReLU. The Conv2D utilized within these layers a 3×3 convolution layer with a stride of 1. The 2D TransposeConv consists of a 2×2 transpose convolution layer with a stride of 2, upsampling the resolution back to 1/8.

2. Qualitative Results

We present qualitative results obtained from our experiments. Fig. 3 illustrates both low-level and high-level BEV features obtained from RadarDistill. The images in the first and third columns in Fig. 3 represent features derived from the LiDAR and radar branches of RadarDistill, respectively. Additionally, to assess the impact of knowledge distillation, we visualize features obtained from the radar-based baseline model (PillarNet-18), shown in the second column of Fig. 3. It is evident that RadarDistill successfully produces radar features that closely resemble those derived from LiDAR. We emphasize the significant improvements achieved by RadarDistill by highlighting key areas using red and blue boxes.

Fig. 4 compares the object detection results in BEV between the conventional method and RadarDistill. Additionally, we provide visualizations of both low-level and high-level features. Notice that several false positives generated by the conventional method are effectively fixed by our RadarDistill. Fig. 5 further demonstrates that RadarDistill is capable of correcting false negatives that arise from the conventional method.

*Equal contributions
†Corresponding author

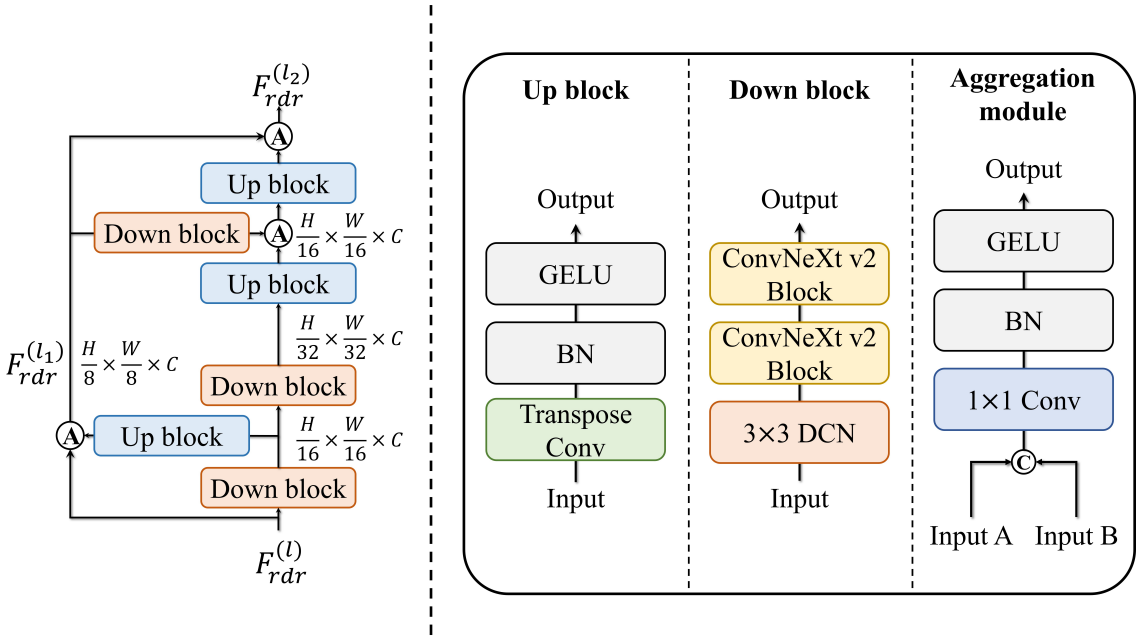

Figure 1. **Detailed structure of CMA module.**

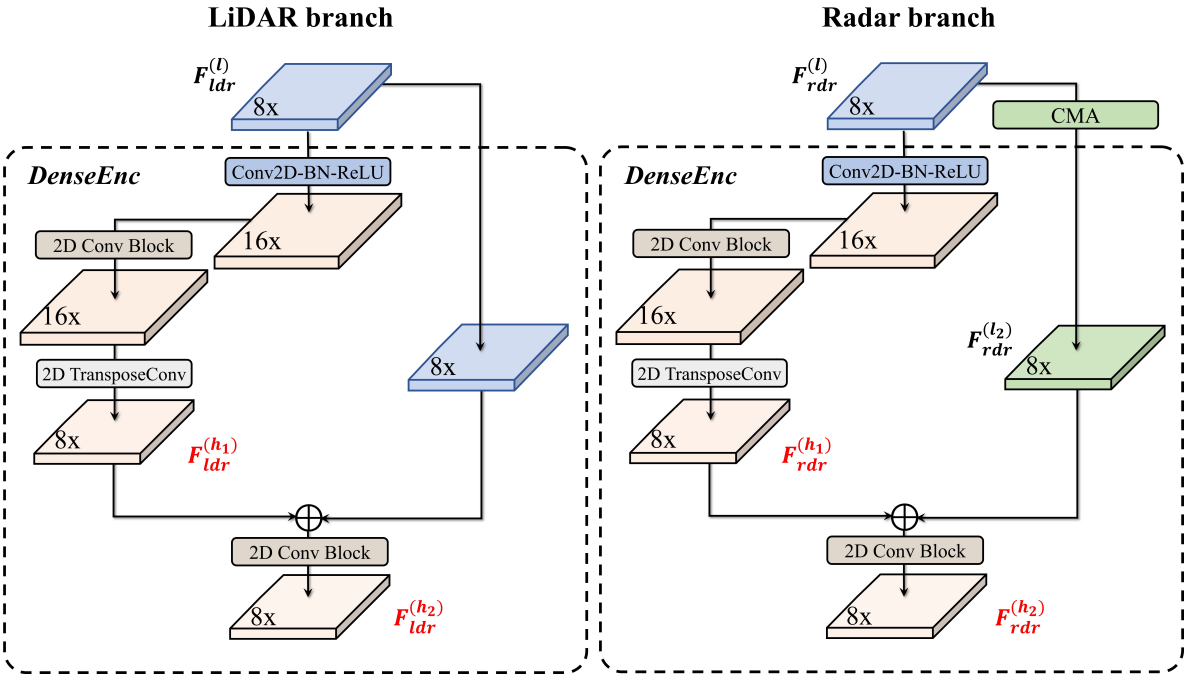

Figure 2. **Detailed structure of DenseEnc.**

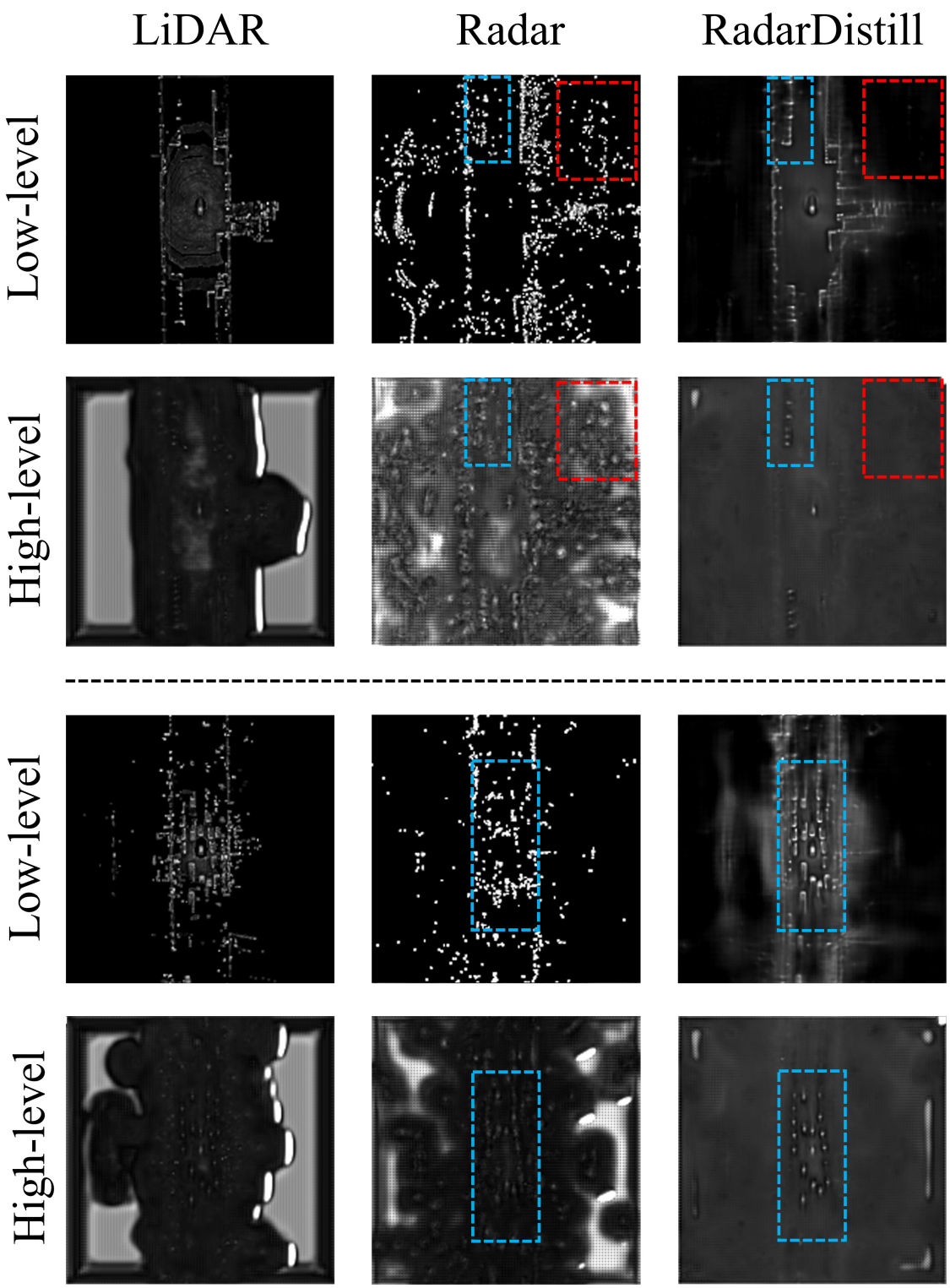

Figure 3. **Feature visualization:** The images in the first and third columns depict features obtained from the LiDAR and radar branches of RadarDistill, respectively. Meanwhile, the images in the second column represent features obtained from the conventional radar-based detector.

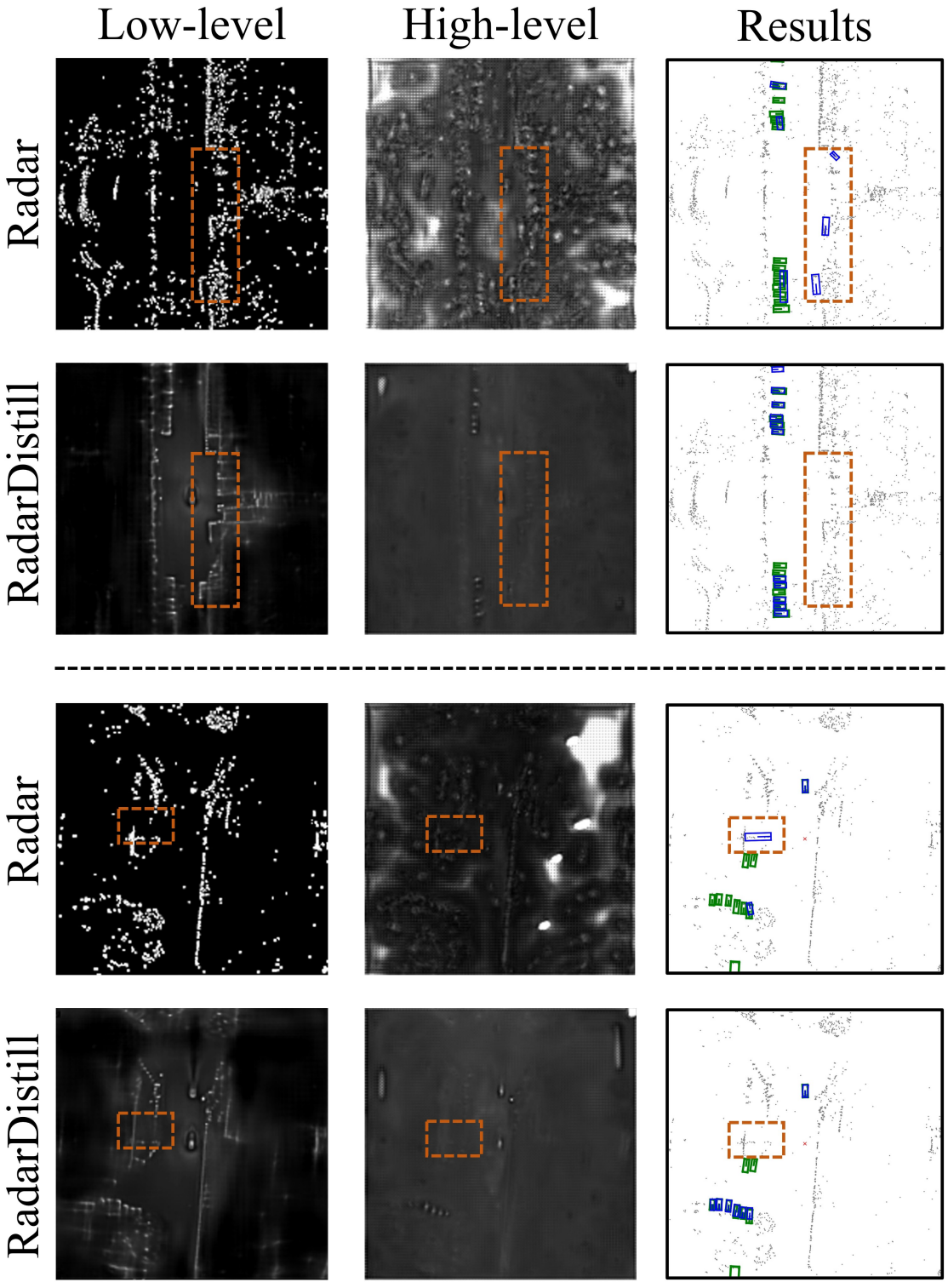

Figure 4. **Feature visualization and detection results:** The green boxes represent the ground truth boxes, while the blue boxes represent the detected boxes. The orange boxes highlight areas where mis-detections from the radar-based model are corrected via knowledge distillation.

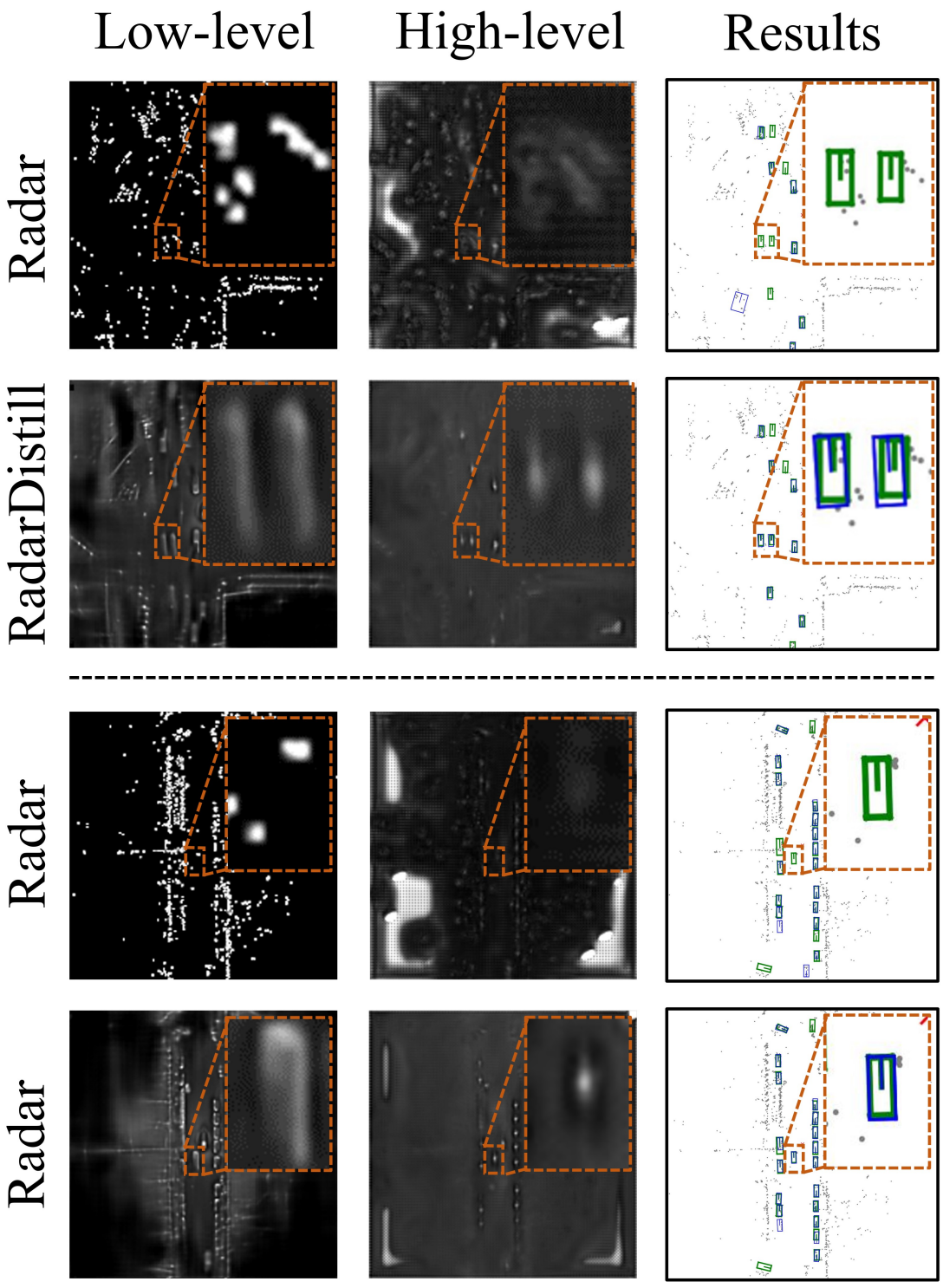

Figure 5. **Feature visualization and detection results:** The green boxes represent the ground truth boxes, while the blue boxes represent the detected boxes. The orange boxes highlight areas where mis-detections from the radar-based model are corrected via knowledge distillation.

References

[1] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 764–773, 2017. 1

[2] Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16133–16142, 2023. 1