

Computer vision researchers develop bilateral reference framework for high-resolution dichotomous image segmentation

August 22 2024



Visual comparison between the results of our proposed BiRefNet and the latest state-of-the-art methods (e.g., IS-Net and UDUN) for high-resolution dichotomous image segmentation (DIS). Details of segmentation are zoomed in for better display. Credit: *CAAI Artificial Intelligence Research* (2024). DOI: 10.26599/AIR.2024.9150038

A research team has developed a computer vision technique that can perform dichotomous image segmentation, high-resolution salient object detection, and concealed object detection in the same framework. Their novel bilateral reference framework (BiRefNet) is able to capture tiny-pixel features and holds potential for a wide range of practical computer vision applications.

The work is published in the journal [CAAI Artificial Intelligence Research](#).

In computer vision research, image [segmentation](#) technology involves separating [digital images](#) into meaningful parts. Through this process, images are easier to analyze. As high-resolution image acquisition has advanced, scientists are now able to achieve highly precise object segmentation.

This new technology is called high-resolution dichotomous image segmentation (DIS), and companies such as Samsung, Adobe, and Disney are now using it. However, current strategies used in DIS are not sufficient to capture the very finest features. To meet these existing challenges in high-resolution DIS, the research team has developed a bilateral reference module.

The team achieved high-resolution DIS with high accuracy through their BiRefNet. "With the proposed bilateral reference module, BiRefNet shows much higher precision on high-resolution images, especially those with fine details. Our BiRefNet is, so far, the best open-source and commercially available model for foreground object extraction," said Deng-Ping Fan, a professor at Nankai University.

The team's novel progressive bilateral reference network BiRefNet handles the high-resolution DIS task with separate localization and reconstruction modules. For the localization module, they extracted hierarchical features from the vision transformer backbone, which are then combined and squeezed. For the reconstruction module, they further designed the inward and outward references as bilateral references, in which the source image and the gradient map are fed into the decoder at different stages.

Instead of resizing the original images to lower-resolution versions to ensure consistency with decoding features at each stage, they kept the original resolution for intact detail features in inward reference and adaptively cropped them into patches for compatibility with decoding

features.

Their BiRefNet provides a simple yet strong baseline that performs high-quality DIS. Its inward reference with source image guidance fills in the missing information in the fine parts and its outward reference with gradient supervision allows it to focus more on regions with richer details.

Because of its extremely accurate segmentation results, BiRefNet has many useful applications. It can be employed in scenarios that common segmentation models cannot handle. For instance, it can accurately find cracks in walls, help maintain them, and determine when to repair them. It can also achieve highly accurate extraction of objects with fine grids and dense holes.

BiRefNet has already been widely used in the computer vision community. It has been integrated into the web app ComfyUI system as the so far best image matting node for better stable-diffusion-based image synthesis. BiRefNet is also widely used for human or portrait segmentation in both images and videos.

Looking ahead, the team plans to extend BiRefNet to more related tasks, including DIS, high-resolution salient object detection, camouflaged object detection, portrait segmentation, and prompt-guided object extraction. The team has already provided well-trained models for most of the aforementioned tasks.

They are also working to adapt BiRefNet to a more lightweight architecture for faster inference on high-resolution images and easier deployment on edge devices. "We have already provided BiRefNet in different parameter magnitudes, some of which have achieved 30 frames per second on images in 1024 x 1024 resolution," said Fan.

"The ultimate goal is to keep our BiRefNet as the best open-source model for a series of related tasks, such as foreground object extraction, image matting, and portrait segmentation, making it strong, free, and open-source forever for everyone," said Fan.

More information: Peng Zheng et al, Bilateral Reference for High-Resolution Dichotomous Image Segmentation, *CAAI Artificial Intelligence Research* (2024). [DOI: 10.26599/AIR.2024.9150038](https://doi.org/10.26599/AIR.2024.9150038)

Provided by Tsinghua University Press

Citation: Computer vision researchers develop bilateral reference framework for high-resolution dichotomous image segmentation (2024, August 22) retrieved 23 August 2024 from <https://techxplore.com/news/2024-08-vision-bilateral-framework-high-resolution.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.