Classification of Biomedical Journal Images using Retargeting-Based Data Augmentation and Visually Explainable Attention Priors

Vinit Veerendraveer Singh and Chandra Kambhamettu

University of Delaware

Abstract

Convolutional Neural Network (CNN)-compatible retargeting-based data augmentations attend to task-critical regions in images and enhance their spatial coverage. This paper proposes a data augmentation approach that utilizes visual explanation techniques to produce attention maps for CNN-compatible retargeting-based data augmentations.

Introduction

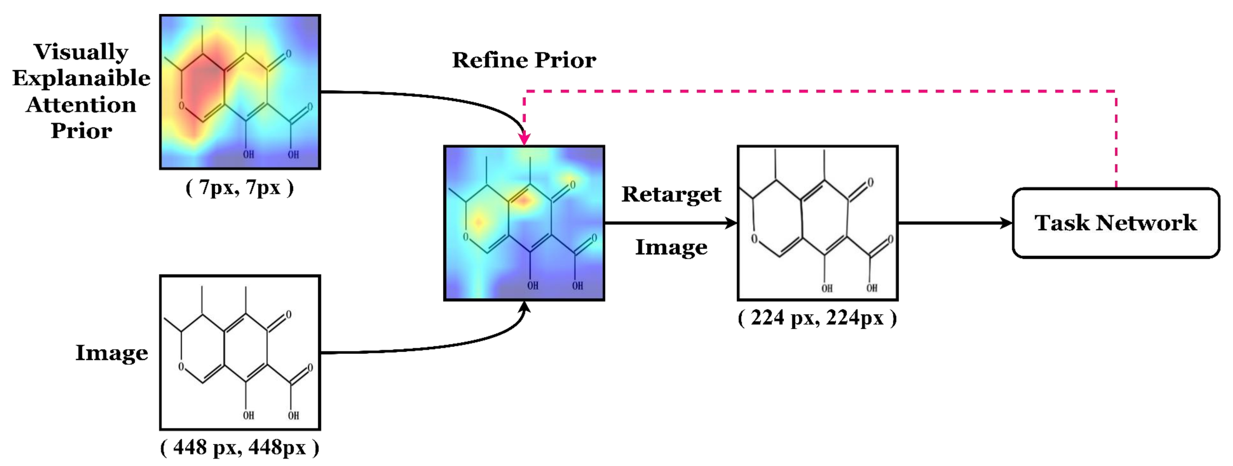

Fig. 1: Visual illustration of our approach

We introduce a novel end-to-end CNN-compatible retargeting-based image augmentation strategy. The task-critical image regions for retargeting are inferred by techniques that produce visual explanations for a CNN's decisions. Thus, compared to all the existing retargeting-based data augmentation approaches, our approach reasonably conserves the original CNN architecture. In addition, the search space to learn the attention priors for image retargeting is reduced.

Method

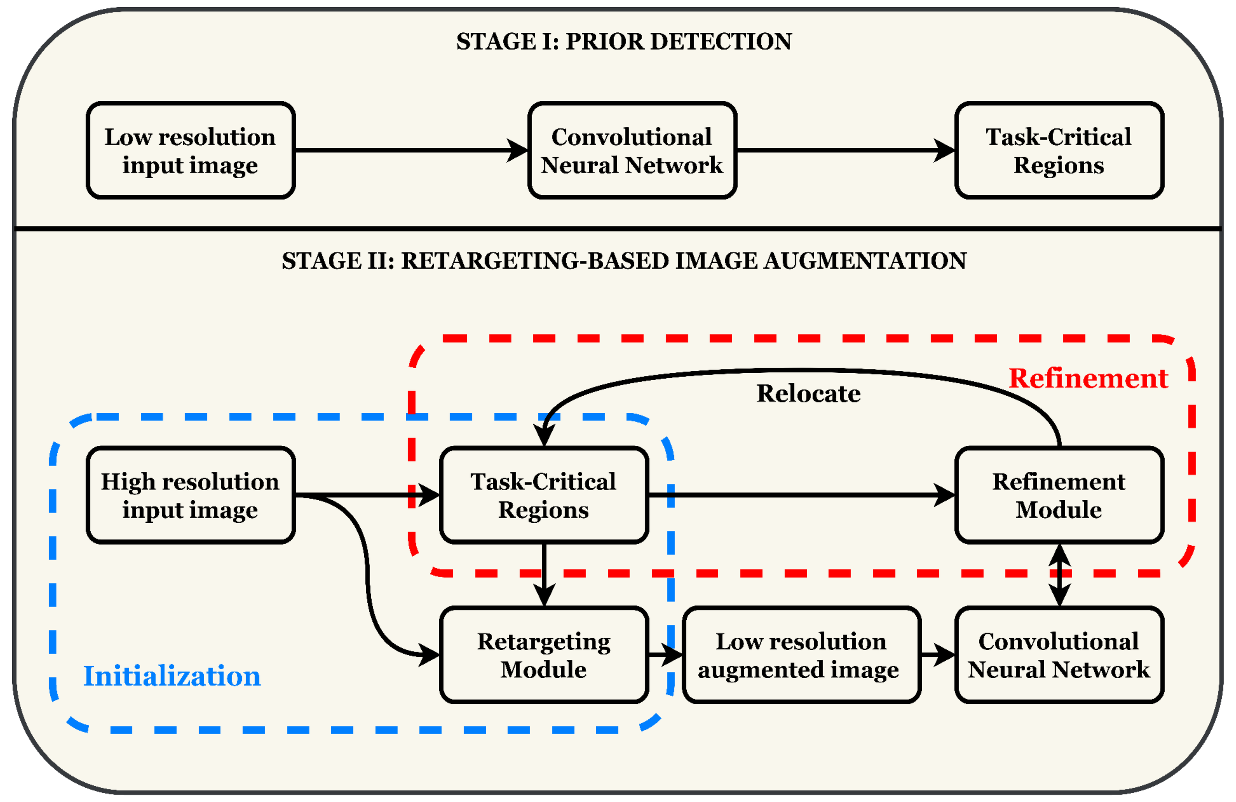

Fig. 2: Overview of our approach

Our data augmentation approach consists of two stages.

In Stage I, a CNN is trained on low-resolution images. Then, techniques that produce visual explanations for the network's decisions output the task-critical regions. We consider these visual explanations as prior knowledge for Stage II.

In Stage II, these priors are employed to perform retargeting-based image augmentations where the spatial coverage of task-critical regions is increased. The second stage is further categorized into two sub-categories:

- **Initialization mode**
- **Refinement mode**

Initialization Mode

In the initialization mode, higher-resolution images are sent to the retargeting module along with the prior. Then, the retargeting module outputs images at a lower resolution, but the spatial coverage of task-critical regions is increased or at least conserved.

Refinement Mode

In the refinement mode, a refinement module jointly optimizes with the network to relocate task-critical regions in the priors acquired from Stage I.

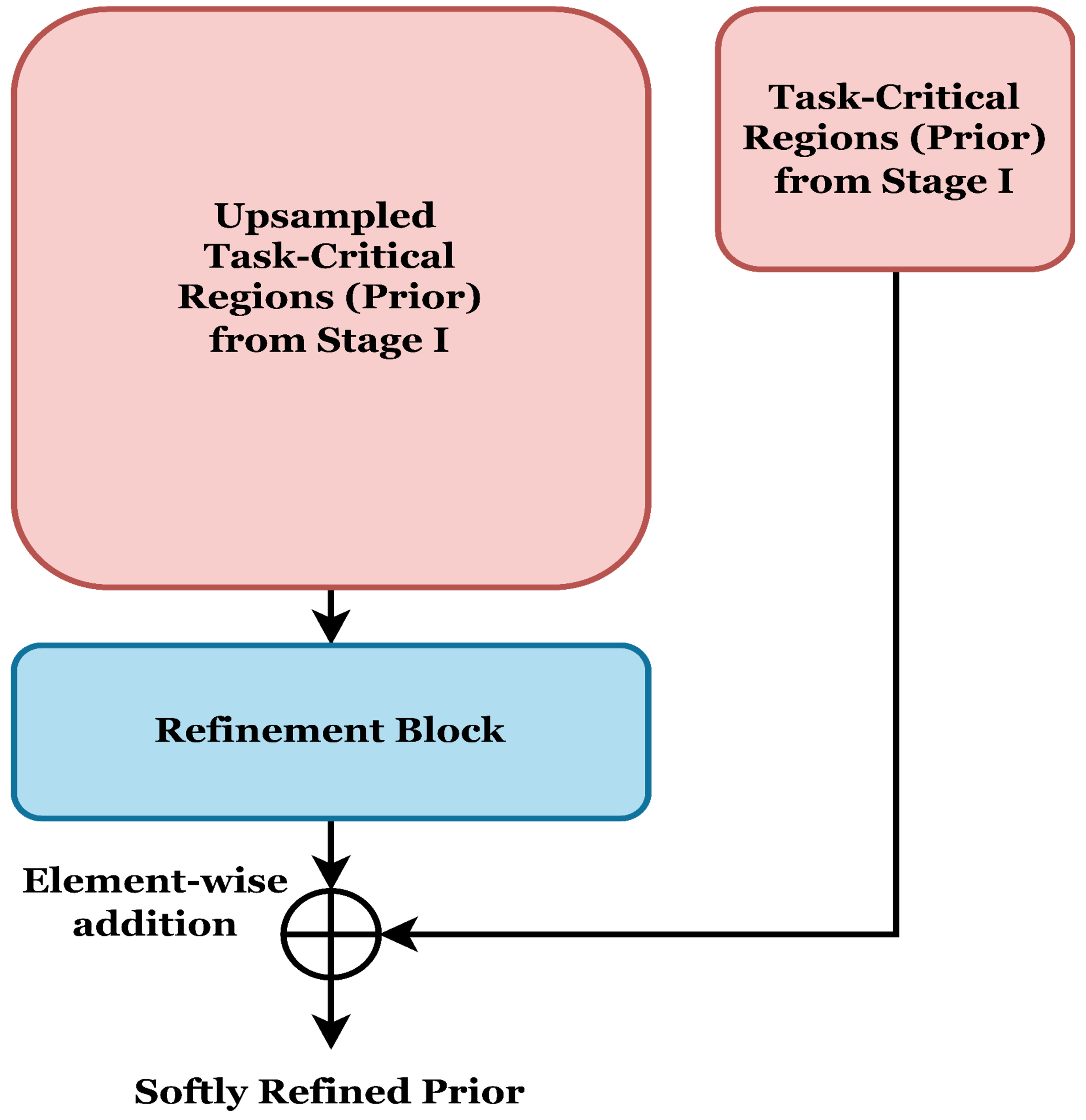

Fig. 3: Refinement Module Architecture

Results and Conclusion

- We evaluated our approach to categorize images in the datasets containing images obtained from biomedical journals
- Augmentation approaches that have demonstrated excellent results on images of natural scenes did not perform well for classifying biomedical document images. On the ResNet-50 and DenseNet-121 models, our approach outperformed seven state-of-the-art augmentation approaches on the ImageCLEF2013, ImageCLEF2015, and ImageCLEF2016 datasets
- In our approach, since attention is inferred by visual explanation methods near the final convolutional layers of the CNN, the size of the localization map is small. As a result of the small localization map, we were able to reduce the spatial complexity of an existing Retargeting Module from $O(N)$ to $O((\log N)^2)$
- Further work will be necessary to demonstrate that our approach generalizes to datasets from other domains and other tasks