Mask-Aware Cropping: Mitigating Mask Imbalance in Segmentation Tasks

Robin Ghyselinck[*], Valentin Delchevalerie[†], Benoît Frénay, Bruno Dumas

University of Namur - NaDI - Faculty of Computer Science
Rue Grangagnage, 21, 5000 Namur - Belgium

Abstract. Data imbalance can take various forms, such as uneven class distributions in the dataset. Solutions like data augmentation, sampling techniques and weighted loss functions are commonly used to address this issue. However, in segmentation tasks, an additional type of imbalance may occur at the pixel-level, with most of them belonging to the background class. This work introduces Mask-Aware Cropping (MAC), a technique to reduce pixel-level imbalance by cropping image regions containing key information about the minority class.

1 Introduction

Data imbalance is a common issue in deep learning applications [1]. Its most common manifestation stands in an asymmetry in the class distributions where one or more classes are underrepresented in the data. Several solutions exist to mitigate this issue, like data augmentation, data sampling, or the use of a weighted loss function. However, when dealing with segmentation tasks, another form of data imbalance can occur at the pixel-level, where most of them often belong to the background class [2]. This imbalance can lead to models that are biased towards predicting background instead of the object of interest. Furthermore, random cropping [1], a common data augmentation method, can inadvertently exacerbate this pixel-level class imbalance by cropping regions with fewer information regarding the non-background class. This imbalance cannot be solved with common techniques because they do not alter the proportion of mask size relative to the overall image.

To address the challenge of pixel-level imbalance in masks, we propose Mask-Aware Cropping (MAC), a novel data augmentation technique designed to handle class imbalance in binary segmentation tasks by performing pixel oversampling for the minority class. MAC selectively crops specific regions of images to ensure that the cropped image is sufficiently informative regarding the minority class. We validate the effectiveness of MAC by comparing its performance against other state-of-the-art methods to deal with class imbalance. Although this issue is sometimes neglected by practitioners, we show that MAC can improve performance when data is strongly imbalanced, and is otherwise at least able to achieve performances similar to those of other methods.

The rest of the paper is structured as follows. Section 2 presents the related works on dealing with class imbalance. Section 3 details our proposed MAC

[*] Supported by FSR, University of Namur

[†] Supported by SPWR under grant n°2010235 - ARIAC by DIGITALWALLONIA4.AI.

method, while the experimental setup, results and discussion are presented in Section 4. Finally, Section 5 presents the conclusions of this study.

2 Related Works

There are two main ways to mitigate class imbalance regardless of the task that should be performed (classification or segmentation) [1]: (i) adding penalization terms in the loss and (ii) modifying the data distribution using sub/oversampling or data augmentation. The first solution aims at penalizing errors on the minority class by increasing its effect on the loss. One can cite the focal loss [3] or the work of Khan et al. [4] who use a cost-sensitive deep neural network that automatically sets a class-dependent weight based on the distribution of the data. For the second category of solutions, one can cite the Synthetic Minority Over-Sampling TEchnique (SMOTE) [5] and its many variants.

For segmentation tasks in particular, another type of class imbalance, often neglected by practitioners, occurs at the pixel-level. Indeed, for many segmentation tasks, most of the pixels in the dataset belong to the background class. To solve this issue, Kochkarev et al. [6] introduced a technique to deal with imbalanced image multi-class segmentation tasks using probabilistic cropping method. Unfortunately, their proposed testing of the method contains flaws such as evaluation on the training data only. Yet, to allow for a fair comparison with MAC (our method), we included their technique in the experimental setup. Another technique by Zaridis et al. [7] uses a pre-processing by training a model to detect the region of interest in the image, before training a second model on the segmentation task. This approach is cumbersome since it involves training two models, and leaves no control over the cropping strategy.

3 Methodology

We propose Mask-Aware Cropping (MAC), a novel data augmentation technique specifically designed to handle pixel-level class imbalance in binary segmentation by extracting sub-regions of the masks and images (i.e., the crops). With MAC, masks and images are cropped such that at least a certain ratio of the pixels contain the minority class as illustrated in Figure 1. This technique relies on three constraints that are controlled by a set of three hyperparameters.

The first of the three hyperparameters is the minimum ratio $0 \leq \rho \leq 1$ of target class pixels required within a crop to ensure that the crop is informative enough about the minority class. If the original image I has width I_w and height I_h, and $C^{(x,y)}$ defines the mask corresponding to a crop in that image I of size $C_w \times C_h$, $\rho = \frac{1}{C_w C_h} \sum_{i,j=1}^{C_w, C_h} C^{(i,j)}$. The second is the stride σ when sliding a window across the image and mask to test the different possibilities for cropping. The third is the frequency τ at which crops from the MAC-coordinates list (see Algorithm 1) are sampled, balancing between background and target samples.

The MAC Algorithm extracts the MAC-coordinates of all image crops having a minimum ratio of non-background above ρ (i.e., valid pixels), moving the cropping window with a stride of σ. During training, one can randomly sample cropped images among the MAC-coordinates list, at a frequency defined by τ. Random cropping is applied at a frequency $1 - \tau$, or when the image contains no valid pixels.

Algorithm 1 MAC Algorithm

```
Require: $\rho \in [0, 1]$ and $(C_h, C_w) \leq (I_h, I_w)$
Ensure: List of $mac\_coordinates$ $(x, y)$ from valid crops in $I$
function GetMacCoordinates($I, (C_h, C_w), \sigma, \rho$)
    Initialize $mac\_coordinates \leftarrow []$
    $(I_h, I_w) \leftarrow$ shape of $I$
    for $y = 0$ to $I_h - C_h$ step $\sigma$ do
        for $x = 0$ to $I_w - C_w$ step $\sigma$ do
            $crop \leftarrow I[y : y + C_h, x : x + C_w]$
            $ratio \leftarrow \frac{\text{sum}(crop \neq 0)}{\text{size}(crop)}$
            if $ratio \geq \rho$ then
                Append $(x, y)$ to $mac\_coordinates$
            end if
        end for
    end for
    return $mac\_coordinates$
end function
```

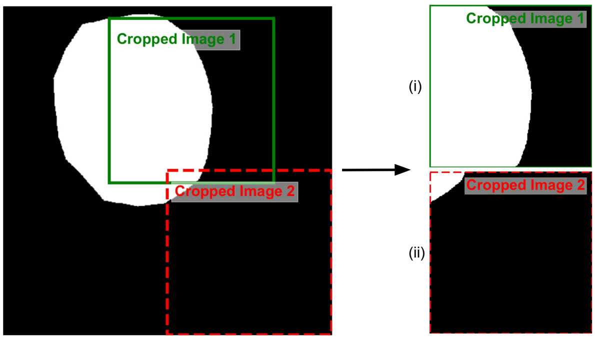

Fig. 1: Example of cropped image masks using random cropping from a gastrointestinal polyp dataset, Kvasir-SEG [8]: (i) is informative about the foreground while (ii) contains few foreground pixels and almost only background pixels.

4 Experimental Setup

This section explains the chosen architecture for all the experiments, then it presents the datasets and hyper-parameters selection, before discussing the considered performance metrics and comparison methods.

4.1 U-Net Architecture

To perform the segmentation tasks and assess the ability of MAC to address class imbalance in segmentation, a standard U-Net [9] with a fixed architecture made of 1,080,929 parameters is used. The U-Net is made of two convolutional layers followed by 4 down sampling (with max pooling) and 4 up scaling (with bilinear upsampling) blocks, and its convolutions use a kernel size of 3×3, are followed by batch normalization layers and ReLU activation functions throughout the network. The first convolutional layer takes 224×224 input images and the down sampling blocks are made of 32, 64, 128, and 256 filters. Finally, a convolutional layer reduces the number of channels to one to perform the binary segmentation.

4.2 Datasets

Three well-known public datasets are used to assess the performance of MAC on binary segmentation. These datasets vary in size and show different levels of pixel-level class imbalance. This allows for a thorough evaluation of how MAC performs across datasets with diverse characteristics in terms of both scale and imbalance. First, URDE [10] is a urban road defects dataset made of $7,000$ images of 1024×1024 pixels. The dataset has only 3.0% of mask pixels containing items of interest. Second, Kvasir-SEG [8] is a gastrointestinal polyp datasets with 1,000 images varying in size from 625×513 pixels to 1920×1072 pixels. On average, 15.8% of the mask pixels contain polyps. Third, HAM10000 [11] is a skin lesion dataset that contains $10,015$ images of 800×600 pixels where 30% of the pixels show a lesion.

4.3 Hyper-parameters

Several configurations are tested for the three datasets. Those configurations are the different combinations of the following choices for the meta-parameters of MAC: $\sigma \in \{16, 32, 64\}$, $\tau \in \{2, 3, 5\}$ and $\rho \in \{0.25, 0.5, 0.75\}$. In addition, for all the three datasets, images are initially resized to 448×448 pixels. The model is trained for 200 epochs with a batch size of 32 using the AdamW optimizer with a learning rate of 0.001, and a reduce on plateau scheduler that divides the learning rate by 3 every 10 epochs without any improvement.

4.4 Performance Assessment

The performance of MAC in binary segmentation is assessed through five performance metrics, namely the Dice Score (DS), Intersection over Union (IoU), precision, recall, and accuracy. This performance is compared against several setups that only differ in how they sample data: (i) No Cropping (NC) is performed on the images, (ii) Random Cropping (RC) as usually done in standard pipelines, and (iii) Probabilistic Cropping (PC), an implementation close to the method by Kochkarev et al. [6] (see Section 2) that draws masks according to a 2D probability density function derived from the 2D distance of the pixels from the background. In addition, the use of a weighted loss function is also considered for both NC (NCW) and RC (RCW) as it is a common solution for dealing with imbalanced data (see Section 2).

5 Experimental Results

The mean metrics from the 5-fold cross-validation are presented in Table 1 for each of the six configurations presented in Section 4 with 95% Confidence Interval (CI). First, the best MAC model uses $\rho = 0.5, \tau = 3$ and $\sigma = 16$ for the URDE dataset. Second, it uses $\rho = 0.25, \tau = 5$ and $\sigma = 16$ for the kvasir-SEG dataset. Third, it uses $\rho = 0.5, \tau = 5$ and $\sigma = 16$ for the HAM10000 dataset. The results show an outperformance by MAC for binary segmentation on Kvasir-SEG and URDE, indicated by higher DS and IoU. All other techniques have much lower DS and IoU. The precision of MAC is either equal or slightly below that of RC, indicating that MAC makes slightly more false positive than RC. Moreover, recall is much higher for RCW and NCW, but those methods yield poor DS, IoU and precision. This means that adding weights to the loss results in over prediction of mask pixels against background ones. For the less imbalanced dataset (30% of its pixels that are not background), HAM10000, DS and IoU show that MAC is comparable but slightly below than RC and PC. However, MAC has a higher precision. This shows that MAC works best for stronger imbalance in the data, and has comparable results when the imbalance is lower.

6 Conclusion

In this paper, we propose a new cropping technique, Mask-Aware Cropping (MAC) that effectively addresses the class imbalance issue that can occur at the pixel-level in binary segmentation. Indeed, for many segmentation tasks, most pixels in the masks often belong to the background class. MAC is based on the definition of new hyperparameters—ρ, σ, and τ—that control the cropping process, offering flexibility over the data augmentation pipeline. The performance of MAC is assessed on three datasets featuring different levels of imbalance, image size and nature. Out of the three, its outperformance is demonstrated on two most imbalanced datasets, URDE [10] and kvasir-SEG [8], with 3% and 16% of the pixels representing the foreground, respectively. On the less imbalanced dataset, HAM10000 [11] (30% pixel-level imbalance), MAC achieves similar performance than the best performing model. Future work will expand MAC to semantic and instance segmentation tasks with more than one class.

References

[1] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *J. Big Data*, 6(1), 2019.

[2] Haibo He and Edwardo A. Garcia. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.*, 21(9):1263–1284, 2009.

[3] Tsung-Yi Lin, Priya Goyal, Ross Girshick, et al. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach.*, 42(2):318–327, 2020.

Table 1: Model performances (95% CI from 5-fold cross-validation)

Model	**DS**	**IoU**	**Precision**	**Recall**	**Accuracy**
URDE: very strong imbalance (3% of pixels from minority class)					
MAC*	**65.5** ± 4.2	**56.3** ± 4.0	44.4 ± 2.1	42.0 ± 3.2	**98.1** ± 0.1
RC	63.4 ± 3.9	54.3 ± 4.0	**45.1** ± 3.0	42.0 ± 2.4	**98.1** ± 0.1
PC	56.1 ± 4.9	47.9 ± 5.0	34.3 ± 2.5	31.8 ± 2.5	97.4 ± 0.0
NC	55.4 ± 15.9	47.4 ± 15.8	39.0 ± 4.8	42.3 ± 4.4	98.0 ± 0.5
RCW	34.7 ± 2.1	25.0 ± 1.7	25.0 ± 1.2	60.1 ± 2.9	95.2 ± 0.1
NCW	32.5 ± 6.3	23.3 ± 5.6	21.6 ± 4.0	**62.2** ± 2.4	94.4 ± 1.1
Kvasir-SEG: strong imbalance (16% of pixels from minority class)					
MAC*	**75.9** ± 2.3	**61.7** ± 3.1	**81.4** ± 3.2	72.5 ± 4.1	**92.9** ± 0.7
RC	74.4 ± 2.5	59.9 ± 3.0	**81.4** ± 2.7	70.0 ± 5.1	92.6 ± 0.5
PC	66.0 ± 7.6	50.2 ± 9.0	65.8 ± 10.1	68.3 ± 5.0	89.2 ± 2.7
NC	66.7 ± 3.8	50.7 ± 4.3	68.4 ± 6.6	66.9 ± 3.0	89.7 ± 1.6
RCW	65.9 ± 5.9	49.8 ± 6.5	54.4 ± 7.6	**86.3** ± 2.6	86.0 ± 4.1
NCW	60.4 ± 5.9	43.9 ± 6.2	50.4 ± 10.2	79.9 ± 8.5	83.8 ± 4.3
HAM10000: moderate imbalance (30% of pixels from minority class)					
MAC*	88.9 ± 1.1	80.4 ± 1.8	**88.4** ± 2.9	89.9 ± 1.8	94.0 ± 0.7
RC	**89.3** ± 0.6	**81.0** ± 0.9	88.2 ± 0.5	90.9 ± 1.5	**94.2** ± 0.2
PC	89.1 ± 1.2	80.8 ± 2.0	87.8 ± 2.6	91.1 ± 0.6	94.1 ± 0.8
NC	88.6 ± 0.7	79.9 ± 1.1	86.8 ± 1.4	91.0 ± 1.2	93.8 ± 0.4
NCW	87.7 ± 0.8	78.5 ± 1.1	84.0 ± 3.1	92.5 ± 4.9	93.2 ± 0.2
RCW	84.9 ± 1.8	74.4 ± 2.6	77.0 ± 3.2	**95.8** ± 0.6	90.9 ± 1.1

*Only best performing models are kept for MAC

[4] Salman H. Khan, Munawar Hayat, Mohammed Bennamoun, et al. Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE Trans. Neural Netw. Learn. Syst.*, 29(8):3573–3587, 2018.

[5] N V Chawla, K W Bowyer, L O Hall, and W P Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.*, 16:321–357, 2002.

[6] Alexey Kochkarev, Alexander Khvostikov, Dmitry Korshunov, et al. Data balancing method for training segmentation neural networks. *GraphiCon*, 2020.

[7] Mylona Zaridis, Dimitris et al. A smart cropping pipeline to improve prostate's peripheral zone segmentation on mri using deep learning. *EAI ETBB*, 1(4):173546, April 2022.

[8] Debesh Jha, Pia H Smedsrud, Michael A Riegler, et al. Kvasir-seg: A segmented polyp dataset. In *MMM*, pages 451–462, 2020.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241, 2015.

[10] Asanka De Silva, Rajitha Ranasinghe, Arooran Sounthararajah, et al. A benchmark dataset for binary segmentation and quantification of dust emissions from unsealed roads. *Sci. Data*, 10(1):14, 2023.

[11] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data*, 5(1):180161, 2018.