Lecture Notes in Networks and Systems 1013

Xin-She Yang
Simon Sherratt
Nilanjan Dey
Amit Joshi *Editors*

Proceedings of Ninth International Congress on Information and Communication Technology

ICICT 2024, London, Volume 3

Springer

Lecture Notes in Networks and Systems

Volume 1013

The series "Lecture Notes in Networks and Systems" publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Xin-She Yang · Simon Sherratt · Nilanjan Dey ·
Amit Joshi
Editors

Proceedings of Ninth International Congress on Information and Communication Technology

ICICT 2024, London, Volume 3

Springer

Editors
Xin-She Yang
Middlesex University
London, UK

Simon Sherratt
University of Reading
Reading, UK

Nilanjan Dey
Department of Computer Science
and Engineering
Techno International New Town
Kolkata, West Bengal, India

Amit Joshi
Global Knowledge Research Foundation
Ahmedabad, Gujarat, India

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-981-97-3558-7 ISBN 978-981-97-3559-4 (eBook)
https://doi.org/10.1007/978-981-97-3559-4

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Addressing Class Imbalance Problem in Semantic Segmentation Using Binary Focal Loss

Rushikesh Chopade, Aditya Stanam, and Shrikant Pawar

Abstract Image segmentation is a foundational technique in computer vision with wide-ranging applications, including its critical role in medical imaging for object identification, automatic labeling, and disease diagnosis. Advancements in deep learning have significantly improved the accuracy and efficiency of image segmentation, making it an increasingly valuable tool in various domains. Class imbalanced datasets are a frequent problem experienced when trying to train segmentation networks. Class imbalance occurs when some classes (semantic categories) in the image have significantly more instances (pixels) than others. In semantic segmentation, this often happens because certain object categories are more prevalent in the real world or dataset, while others are rarer. When training a deep learning model for semantic segmentation, this imbalance can lead to several problems. In this article, we have experimented with the class weightage parameters of binary focal loss to address the class imbalance problem in semantic segmentation. By utilizing the CANDID-PTX dataset, we have utilized U-Net architecture containing upsampling (encoder) and a downsampling (decoder) network for comparing binary focal loss rates among different alpha and gamma coefficients class weights. Doing so, we found that the adjustment of class weights in the loss function could notably help in resolving the class imbalance problems.

Keywords Image segmentation · Binary focal loss · Class imbalance

R. Chopade
Department of Geology and Geophysics, Indian Institute of Technology, Kharagpur, Kharagpur, West Bengal 721302, India

A. Stanam
University of Iowa, Iowa City, IA 52242-5000, USA
e-mail: aditya-stanam@uiowa.edu

S. Pawar (✉)
Department of Computer Science and Biology, Claflin University, Orangeburg, SC 29115, USA
e-mail: spawar@claflin.edu

X.-S. Yang et al. (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, Lecture Notes in Networks and Systems 1013,
https://doi.org/10.1007/978-981-97-3559-4_28

1 Introduction

Class imbalance is a common challenge in semantic segmentation, and it can degrade the overall performance of the network [1]. Class imbalance occurs when some classes (semantic categories) in the image have significantly more instances (pixels) than others. In semantic segmentation, this often happens because certain object categories are more prevalent in the real world or dataset, while others are rarer. When training a deep learning model for semantic segmentation, this imbalance can lead to several problems as follows:

A. Biased Training: If one class dominates the dataset, the model may become biased toward that class during training. As a result, it might perform well on the majority class but poorly on the minority classes [2].
B. Reduced Generalization: Models trained on imbalanced data may struggle to generalize to underrepresented classes in new, unseen data, as they have not learned enough about these classes [3].
C. Low Intersection over Union (IoU): IoU is a common metric for evaluating semantic segmentation models. When a class is underrepresented, the model may have a low IoU for that class, indicating poor segmentation quality [4].

To address the issue of class imbalance in semantic segmentation, the following techniques can be employed:

A. Data Augmentation: Augmenting the training dataset by applying transformations like rotation, scaling, and flipping can help balance the distribution of class instances. Augmentations can create additional training examples for underrepresented classes [5].
B. Weighted Loss: Assigning different weights to each class during the loss calculation can give more importance to underrepresented classes. This encourages the model to focus on correctly segmenting these classes [5].
C. Oversampling and Undersampling: Oversampling the minority classes (duplicating or generating more examples) or undersampling the majority class (removing some examples) can balance class distribution. These methods should be used judiciously to avoid overfitting or losing valuable data [5].
D. Class-Rebalancing Loss Functions: Some loss functions are designed to address class imbalance explicitly, like the focal loss or class-balanced loss. These loss functions downweight the contribution of well-classified examples and emphasize the importance of hard-to-classify examples [6].
E. Data Stratification: When splitting the dataset into training, validation, and test sets, ensure that each set has a representative distribution of class instances. This helps in evaluating the model's performance on underrepresented classes [6].
F. Synthetic Data Generation: Generating synthetic data for underrepresented classes can help balance the dataset. Techniques like generative adversarial networks (GANs) or data synthesis methods can be employed for this purpose [7].

G. Post-processing: Applying post-processing techniques like conditional random fields (CRFs) can refine the segmentation results and improve the boundaries between classes [8].

In this article, we have tried to address the imbalance problem where the fraction of images containing a certain class is very small in comparison with other classes in the same set using rib fracture labels. The area covered by the positive class or the region of interest (ROI) in the semantic segmentation algorithm affects the overall prediction quality. Evaluation metrics like accuracy do not perform well when the ROI is small. In medical images, small objects in acute bone fractures in X-rays, brain cancer images [9], and red blood cells in peripheral blood smear (PBS) images [10] can be hard to locate using accuracy metric. We have utilized the CANDID-PTX dataset [11] with a positive class (rib fracture mask) in a range of just 0.08–0.5% of the total image area. To address this imbalance problem, we have experimented with the loss function and evaluation metrics. Metrics that focus on only true positive classification without a true negative inclusion provide better performance representation [9]. This is why the dice similarity coefficients (DSC) and IoU are highly popular and recommended metrics in the field of medical image segmentation [9].

2 Method

2.1 Dataset

Digital Imaging and Communications in Medicine (DICOM) format chest radiograph images from CANDID-PTX dataset are utilized in this study [11]. It contains a total of 19,237 images, of which 335 images containing acute rib fractures have been used in this study. An acute rib fracture was defined as any rib with cortical disruption visible on a chest radiograph without evidence of healing such as callus formation. There are a total of 973 different annotations provided by different radiologists for these 335 images. The same image having different annotations by different radiologists is treated as different images for dataset enhancement.

2.2 Preprocessing

All the images have 1024 * 1024 pixel resolution with three channels (RGB). The 335 acute rib fracture images have been provided with a run-length encoding notation. Run-length encoding is a simple Morse-like representation of a 2D image [12]. The 1024 * 1024 image is represented as a one-dimensional array with rows appended one after the other. The run-length encoding provided for the images with the dataset contains a string of comma-separated numeric values. When the mask begins, the first number in the string is the pixel number. The numbers hence following are the lengths of the mask and background pixels. Such masks have been derived from the RLE strings for all 973 annotations (replication code link provided in the supplementary section). The maximum pixel intensity in each image is different which can cause problems while training the algorithm. So the pixel intensity of every image has been normalized in a range of 0–1. The masks formed from RLE are 2D and have been converted to 3D to be compatible with the U-Net architecture. Finally, the original image and the masked image have been downsampled to 512 * 512 before feeding to the U-Net algorithm.

2.3 Evaluation Metrics

Various metrics and losses such as IoU, DSC, sensitivity, specificity, accuracy/ rand index, receiver operating characteristics (ROC), area under ROC curve (AUC), Cohen's kappa, average Hausdorff distance, dice loss, Jaccard loss, binary cross-entropy, categorical cross-entropy loss, binary focal loss, categorical focal loss, and dual focal loss have been used to train the semantic segmentation algorithms [9, 13]. Weighted cross and dual cross-entropy can introduce a vanishing gradient, penalize negative classes inversely, or can generate a sub-optimal loss weighting between classes, which limits their ability to improve classification accuracy or ease of use. Focal loss has been proven to be effective at loss balancing by intensifying the loss on hard-to-classify classes [14]. In this study, we have utilized pixel-wise recall as the evaluation metric to measure the model performance. The recall metric will give higher weightage to the positive class as it calculates the fraction of the actual positive class that is predicted positive. Following is the notation utilized for recall:

$$\text{Recall} = \text{True Positives (TP)} \big/ (\text{True Positives (TP)} + \text{False Negatives (FN)})$$

The binary focal loss function is applied using following notation:

$$L\big(y, p^{\wedge}\big) = -\alpha y\big(1 - p^{\wedge}\big)\gamma \log\big({}^{\wedge}p\big) - (1 - y)^{\wedge}p\gamma \log\big(1 - p^{\wedge}\big)$$

where y is ground truth pixel value; $p^{\wedge}$ is predicted pixel value; alpha (α) is the weighting factor which governs the tradeoff between the precision and recall by weighting errors for positive class; and gamma (γ) is focusing parameter for modulating factor specifying how much higher-confidence correct predictions contribute to overall loss.

2.4 Model Architecture

A transformer-based U-Net model containing an upsampling network (encoder) and a downsampling network (decoder) bridged by double convolutional layers is utilized for training [15]. The detailed architecture is attached as an image in the GitHub link provided in the Supplementary File section. The rib fracture dataset is split into 70–30% train-test split for the model training and validation. The model has been trained using a mini-batch gradient optimization algorithm in batches of 4 using a custom image data generator. The algorithm is trained for 10 epochs using Adam optimizer. The learning rate used in the optimizer is 0.001. Finally, the abovementioned binary focal loss with recall evaluation metric has been used to train and validate the algorithm.

3 Results

The training results for the binary focal loss while adjusting the values of alpha and gamma hyperparameters are provided in Table 1. The optimal gradual decay for loss values was found when the alpha and gamma were tuned to 0.01 and 0.1, respectively. The training loss was found to decrease continuously from a maximum value of 0.63–0.05, whereas the validation loss was initially found to be very low (0.18) subsequently increasing to a value of 5. With further epoch progression, it then decreased to a value of 0.32 (Fig. 1).

Table 1 Binary loss progression over 10 epochs when trained with various values of hyperparameters alpha and gamma

Alpha (*default)	Gamma (*default)	Training loss	Validation loss
0.25*	2	0.1 to 0.005	0.06 to 0.006
0.5	2*	0.026 to 0.009	0.017 to 0.016
1	2*	$1e-5$ to $2e-6$	$1e-5$ to ~0
10	2*	−9.2 to −24	−142 to −143.7
0.1	2*	0.19 to 0.002	0.008 to 0.001
0.01	2*	0.46 to 0.008	0.005 to 0.003
0.001	2*	0.1 to 0.001	6 to 0.01
0.01	5	0.013 to $6e-4$	0.012 to $2e-4$
0.01	10	0.001 to $1e-4$	$8e-5$ to $9e-5$
0.01	1	0.26 to 0.01	0.043 to 0.002
0.01	0.1	0.63 to 0.05	5.1 to 0.18
0.01	0.01	0.5 to 0.04	0.17 to 0.019

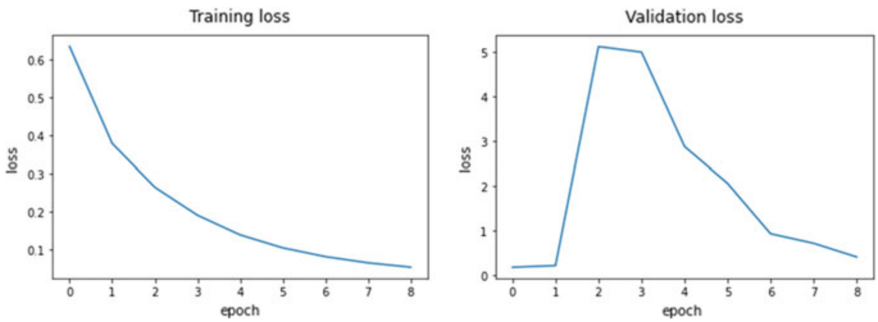

Fig. 1 Training (left) and validation losses (right) for the semantic segmentation algorithm. The final recall value was found to be 0.32 in the last epoch

4 Discussion and Conclusion

An adjustment of class weights in the loss function of the semantic segmentation algorithm could help in resolving the class imbalance problems. Negative losses can be observed when the class weightage to the positive class is not structured properly. The tuning of hyperparameters alpha and gamma in the binary focal loss function can significantly help in addressing the class imbalance problem. In summary, addressing class imbalance is crucial for improving the performance of semantic segmentation models. Employing a combination of these strategies can help mitigate the effects of class imbalance and lead to more accurate and balanced semantic segmentation results, especially for datasets with unequal class distributions.

Author Contributions Rushikesh Chopade, Shrikant Pawar, and Aditya Stanam conceived the concepts, planned, and designed the article. Shrikant Pawar and Rushikesh Chopade primarily wrote and edited the manuscript.

Competing Interests The authors declare that they have no competing interests.

Funding Source This work was primarily supported by the National Science Foundation EPSCoR Program under NSF Award # OIA-2242812.

Supplementary File Training code can be found here: https://github.com/rushikeshchopaderc/Semantic_Segmentation_Code.

Supplementary file 1: Transformer-based U-Net model architecture.

References

1. Megahed FM, Chen YJ, Megahed A, Ong Y, Altman N, Krzywinski M (2021) The class imbalance problem. Nat Methods 18(11):1270–1272. https://doi.org/10.1038/s41592-021-01302-4
2. Raghuwanshi BS, Shukla S (2018) Class-specific extreme learning machine for handling binary class imbalance problem. Neur Netw Official J Int Neur Netw Soc 105:206–217. https://doi.org/10.1016/j.neunet.2018.05.011

3. Desuky AS, Hussain S (2021) An improved hybrid approach for handling class imbalance problem. Arab J Sci Eng 46(4):3853–3864. https://doi.org/10.1007/s13369-021-05347-7
4. Wang D, Wu H (2021) IoU regression with H+L-sampling for accurate detection confidence. Sensors (Basel, Switzerland) 21(13):4433. https://doi.org/10.3390/s21134433
5. Al-Stouhi S, Reddy CK (2016) Transfer learning for class imbalance problems with inadequate data. Knowl Inf Syst 48(1):201–228. https://doi.org/10.1007/s10115-015-0870-3
6. Zhang W, Chen Y, Yang W, Wang G, Xue JH, Liao Q (2021) Class-variant margin normalized softmax loss for deep face recognition. IEEE Trans Neur Netw Learn Syst 32(10):4742–4747. https://doi.org/10.1109/TNNLS.2020.3017528
7. Aloo R, Mutoh A, Moriyama K, Matsui T, Inuzuka N (2022) Ensemble method using real images, metadata and synthetic images for control of class imbalance in classification. Artif Life Robot 27(4):796–803. https://doi.org/10.1007/s10015-022-00781-8
8. Feng LX, Li X, Wang HY, Zheng WY, Zhang YQ, Gao DR, Wang MQ (2021) Automatic sleep staging algorithm based on time attention mechanism. Front Hum Neurosci 15:692054. https://doi.org/10.3389/fnhum.2021.692054
9. Müller D, Soto-Rey I, Kramer F (2022) Towards a guideline for evaluation metrics in medical image segmentation. BMC Res Notes 15(1):210. https://doi.org/10.1186/s13104-022-06096-y
10. Tyas DA, Ratnaningsih T, Harjoko A, Hartati S (2022) Erythrocyte (red blood cell) dataset in thalassemia case. Data Brief 41:107886. https://doi.org/10.1016/j.dib.2022.107886
11. Feng S, Azzollini D, Kim JS, Jin CK, Gordon SP, Yeoh J, Kim E, Han M, Lee A, Patel A, Wu J, Urschler M, Fong A, Simmers C, Tarr GP, Barnard S, Wilson B (2021) Curation of the CANDID-PTX dataset with free-text reports. Radiol Artif Intell 3(6):e210136. https://doi.org/10.1148/ryai.2021210136
12. Sarhan A (2021) Run length encoding based wavelet features for COVID-19 detection in X-rays. BJR Open 3(1):20200028. https://doi.org/10.1259/bjro.20200028
13. Semantic segmentation losses and metrics (2024). Retrieved from: https://github.com/qubvel/segmentation_models/blob/master/segmentation_models/losses.py
14. Hossain S, John B, Andrew P (2020) Dual focal loss to address class imbalance in semantic segmentation. Retrieved from: https://users.monash.edu/~app/papers/20_NeuroC_prostate.pdf
15. Yousef R, Khan S, Gupta G, Siddiqui T, Albahlal BM, Alajlan SA, Haq MA (2023) U-Net-based models towards optimal MR brain image segmentation. Diagnostics (Basel, Switzerland) 13(9):1624. https://doi.org/10.3390/diagnostics13091624