GP-CPS: GRADIENT PENALTY CROSS PSEUDO SUPERVISON FOR SEMI-SUPERVISED MEDICAL IMAGE SEGMENTATION

Peng Jin[1*], *Yuxuan Liu*[2], *Taiwei Cui*[3]

[1]School of Sicence, Harbin Institute of Technology, Shenzhen, China
[2]Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Beijing, China
[3]Khoury College of Computer Science, Northeastern University, Boston, USA
[*]Corresponding to: 23S058001@stu.hit.edu.cn

ABSTRACT

Medical imaging has always been constrained due to the challenges in the acquisition process, high costs, low signal-to-noise ratio, and the complexity of biomedical image features. This paper proposes a semi-supervised model called Gradient Penalty Cross Pseudo Supervision (GP-CPS), which is based on the Cross Pseudo Supervision (CPS) model and innovatively introduces the concept of gradient penalty, significantly enhancing the model's performance. To our knowledge, this is the first time the concept of gradient penalty has been applied in the field of image segmentation. Furthermore, this work introduces a new concept of fused cross pseudo supervision to enhance the diversity of training and strengthen the robustness of the model. Using the publicly accessible Kvasir-SEG dataset, proposed model is compared with baselines and advanced models. In all four partitions containing different proportions of unlabeled data, proposed model consistently demonstrates superior performance.

Index Terms— Medical image segmentation, Semi-supervised learning, Gradient penalty, Pseudo labeling

1. INTRODUCTION

Medical image segmentation is a pivotal technology in the field of medical image processing, playing a crucial role in disease diagnosis, treatment planning, surgical navigation, and therapeutic efficacy assessment [1]. It aids doctors in understanding pathology for precise treatment. However, it's challenging due to small datasets limiting deep learning model training. High data acquisition costs and privacy concerns hinder large annotated dataset creation, complicating neural network training.

To overcome this challenge, researchers have explored various semi-supervised learning techniques to leverage unlabeled data to enhance the model's generalization ability. Mainstream methods fall into two categories: self-training and consistency regularization. The self-training (also known as Pseudo labeling) approaches [2, 3] utilize unlabeled images in a supervised-like manner with one-hot pseudo labels generated by the up-to-date optimized model itself. Consistency regularization methods [4, 5, 6] improve the generalizability by enforcing the consistency among the predictions of unlabeled images with perturbations. Among them, CPC [7] and CPS [8] are two effective strategies. CPC improves classification performance by conducting consistency training on imbalanced datasets, while CPS enhances the model's robustness by using the model's own high-confidence predictions as additional annotation information through cross-pseudo labeling. In addition, the CutMix [9] technique generates pseudo-labeled data by copying and pasting segments of unlabeled data in images, further expanding the training set.

In Generative Adversarial Networks (GANs) [10], gradient penalty technology is widely used to improve the stability of model training and the quality of generation. It enhances the model's generalization ability by penalizing the difference between the model's output and the real data distribution. The gradient penalty strategy has been proven to be highly effective, yet it has not been applied to the field of image segmentation to date.

In this study, we propose an innovative neural network architecture called GP-CPS, which ingeniously integrates the cross pseudo labeling strategy of CPS and the gradient penalty mechanism. Unlike the improvement of WGAN [11] by GP-WGAN [12], our goal is not to optimize the consistency between the network output and the real data distribution, but to focus on reducing the interference of the background part on the target part by limiting the size of the relevant gradients. This innovative gradient penalty method effectively enhances the model's segmentation performance. In order to enhance the effectiveness of the gradient penalty, we introduced the fused cross pseudo labeling, drawing inspiration from CutMix. Such labels can amplify the influence of background pixels on the target pixels, thereby intensifying the effect of the gradient penalty. Proposed model has been rigorously evaluated against both baselines and state-of-the-art models using the Kvasir-SEG dataset [13]. The results demonstrate that GP-CPS outperforms its counterparts, thereby underscoring its exceptional performance.

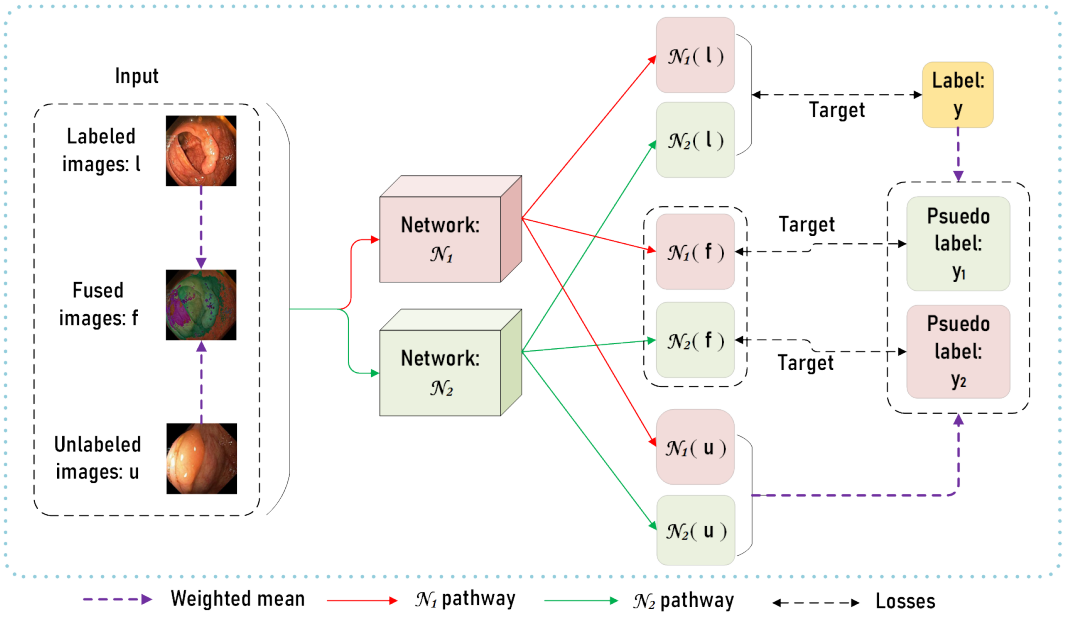

Fig. 1. Overview of the proposed model. Labeled images, unlabeled images, and fused images are all simultaneously input into two neural networks. The labeled images are supervised by the labels, while the pseudo-labels for the fused images are generated by fusing the outputs of the neural networks with the true labels.

2. METHOD

Define the medical image as $m \in \mathbb{R}^{3\times w\times h}$, where w and h represent the width and height of the image, respectively. Our goal is to predict the label map $\hat{y} \in \{0,1\}^{w\times h}$ through semi-supervised learning methods to distinguish the background and target areas in the image. The output of the neural network is $p \in [0,1]^{w\times h}$, consisting of $w \times h$ numbers between 0 and 1.

The dataset is divided into labeled data and unlabeled data, denoted by l and u, respectively. For any labeled image l, and any unlabeled image u, they can be fused into a fused image f through a weighted average, that is

$$f = \varepsilon l + (1-\varepsilon)u \tag{1}$$

where $\varepsilon \sim \text{Uniform}(0,1)$.

Proposed model architecture consists of three main parts: the supervised learning part, the gradient penalty part, and the semi-supervised learning part. The comprehensive structure of the model is shown in Figure 1.

2.1. Gradient Penalty

To theoretically elucidate the rationale behind the gradient penalty trategy, we first consider a medical image $m \in \mathbb{R}^{3\times w\times h}$ and an ideal neural network $\mathcal{N}$. The ideal neural network $\mathcal{N}$ can perfectly distinguish between the background and the target areas in images, that is:

$$\mathcal{N}(x) = \begin{cases} 0, & \text{pixel } x \in \text{background} \\ 1, & \text{pixel } x \in \text{target} \end{cases} \tag{2}$$

Let $\left\{\mathcal{N}^{(1)}, \mathcal{N}^{(2)}, \cdots, \mathcal{N}^{(n)}, \cdots\right\}$ be a sequence of neural networks converging to N, then we have:

$$\lim_{n\to\infty} \mathcal{N}^{(n)} = \mathcal{N} \tag{3}$$

From the image m, we randomly select a pixel x_B that belongs to the background, and a random pixel x_R that is not adjacent to x_B. By inputting the image m into the neural network $\mathcal{N}$, we obtain the output $p = \mathcal{N}(m) \in [0,1]^{w\times h}$. The neural network segmentation results corresponding to the pixels x_B and x_R are denoted as $p_B = \mathcal{N}(x_B) \in [0,1]$ and $p_R = \mathcal{N}(x_R) \in [0,1]$, respectively.

Since $\mathcal{N}$ is an ideal network, it can perfectly identify that x_B belongs to the background, expressed as:

$$\lim_{n\to\infty} \mathcal{N}^{(n)}(x_B) = \mathcal{N}(x_B) = p_B = 0 \tag{4}$$

Let $\delta^{(n)}$ be a lower-order infinitesimal of $\mathcal{N}^{(n)}(x_B)$, such that:

$$\mathcal{N}^{(n)}(x_B) = o(\delta^{(n)}) \to 0 \tag{5}$$

When a perturbation $\delta^{(n)}$ is added to the background pixel x_B, the ideal network $\mathcal{N}$ should still classify the perturbed pixel x'_B as background, meaning the output of $\mathcal{N}^{(n)}(x_B + \delta^{(n)})$ remains an equivalent infinitesimal of $\mathcal{N}^{(n)}(x_B)$:

$$\mathcal{N}^{(n)}(x_B + \delta^{(n)}) \sim \mathcal{N}^{(n)}(x_B) = o(\delta^{(n)}) \to 0 \tag{6}$$

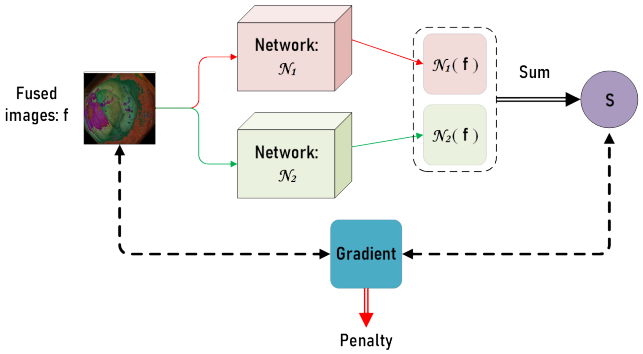

Fig. 2. Gradient Penalty Design in the GP-CPS Model. The fused image f is input into two neural networks, the sum of all elements of the outputs is calculated to obtain S, followed by backpropagation, and finally, the gradient of S with respect to f is constrained.

Since $\delta^{(n)}$ is a lower-order infinitesimal compared to both $\mathcal{N}^{(n)}(x_B)$ and $\mathcal{N}^{(n)}(x_B+\delta^{(n)})$, we have:

$$\begin{aligned}\frac{\partial p_B}{\partial x_B} &= \lim_{\delta\to 0}\frac{\mathcal{N}(x_B+\delta)-\mathcal{N}(x_B)}{\delta}\\ &= \lim_{n\to\infty}\frac{\mathcal{N}^{(n)}(x_B+\delta^{(n)})-\mathcal{N}^{(n)}(x_B)}{\delta^{(n)}}\\ &= 0\end{aligned} \tag{7}$$

Since B and x_R are non-adjacent pixels, the infinitesimal perturbation at the position of x_B will not affect the value at x_R. Thus:

$$\frac{\partial p_R}{\partial x_B} = 0 \tag{8}$$

From Eq. (7) and Eq. (8), it follows that, for the ideal network $\mathcal{N}$, the output gradient with respect to x_B should be zero. To bring the trained neural network closer to this ideal state, we should constrain the network's output gradient concerning the background pixels to approach zero. This process is visually represented in Figure 2. Through the gradient penalty strategy, we can effectively guide the network to learn more stable feature representations.

To implement the gradient penalty, the fused data f is input into neural networks $\mathcal{N}_1$ and $\mathcal{N}_2$, yielding outputs $\mathcal{N}_1(f), \mathcal{N}_2(f) \in [0,1]^{w\times h}$. The sum of all elements of $\mathcal{N}_1(f)$ and $\mathcal{N}_2(f)$ is calculated to obtain the sum of the neural network outputs S:

$$S = \sum_{a\in\mathcal{N}_1(f)} a + \sum_{b\in\mathcal{N}_2(f)} b \tag{9}$$

As previously mentioned, the ideal gradient for S with respect to the background pixels should be zero. By backpropagating S, the gradient G of S with respect to the fused image f is obtained:

$$G = \frac{\partial S}{\partial f} = \frac{\partial\left(\sum_{a\in\mathcal{N}_1(f)} a + \sum_{b\in\mathcal{N}_2(f)} b\right)}{\partial f} \tag{10}$$

Here, the gradient at pixel x is denoted as $G(x)$. Define the set B as the collection of all background pixels in the fused image f:

$$B = \{x | y(x) = 0 \text{ or } \mathcal{N}(x) < 0.5\} \tag{11}$$

Finally, the gradient penalty loss function L_{grad} is obtained by constraining the absolute value of the gradients for all pixels in the set B:

$$L_{grad} = mean_{x\in B}\left(|G(x)|\right) \tag{12}$$

2.2. Fused Cross Pseudo Supervision

To more effectively utilize unlabeled data u, we introduce the concept of fused cross pseudo supervision in proposed model design. The labeled data l and unlabeled data u are averaged with weights to obtain the fused data f. This f is then input into neural networks $\mathcal{N}_1$ and $\mathcal{N}_2$, which produce outputs p_1 and p_2, respectively. Referring to the generation method of the fused data f, fused cross pseudo labels y_1 and y_2 for f are generated from the label y and the neural network outputs p_1 and p_2 as follows:

$$y_1 = \varepsilon y + (1-\varepsilon)p_2 \tag{13}$$

$$y_2 = \varepsilon y + (1-\varepsilon)p_1 \tag{14}$$

where ε is the same as in Eq. (1).

By supervising the fused outputs with the fused pseudo labels, we define the fusion cross-semi-supervised loss function L_{semi}:

$$L_{semi} = L_{dice}(\mathcal{N}_1(f), y_1) + L_{dice}(\mathcal{N}_2(f), y_2) \tag{15}$$

2.3. Loss Function

In the supervised learning component, we input labeled images l into two neural networks $\mathcal{N}_1$ and $\mathcal{N}_2$, obtaining the image segmentation results $\mathcal{N}_1(l)$ and $\mathcal{N}_2(l)$. Subsequently, we calculate the Dice coefficient using the labels y and the segmentation results, thereby obtaining the supervised loss function

$$L_{sup} = L_{dice}(\mathcal{N}_1(l), y) + L_{dice}(\mathcal{N}_2(l), y) \tag{16}$$

This method is simple and effective, providing a solid learning foundation for the model.

The comprehensive training scheme for the neural network depends on three main loss functions: the supervised loss L_{sup}, the gradient penalty loss L_{grad}, and the fusion pseudo-supervision loss L_{semi}. The total loss function is shown as follows:

$$Loss = L_{sup} + \lambda_1 L_{grad} + \lambda_2 L_{semi} \tag{17}$$

3. EXPERIMENTS

3.1. Dataset

This study uses the Kvasir-SEG dataset, which was published in the 2020 MediaEval competition. The dataset contains 1000 images of gastrointestinal polyps. These images were annotated and verified by senior gastrointestinal experts. They also come with segmentation masks. To conduct model training, validation, and testing, we divided these images into three parts: 800 for training, 100 for validation, and 100 for testing. Following the division criteria of the CPC [8] model, we randomly partitioned the training set into two subsets. One subset includes half, one-fourth, one-eighth, and one-sixteenth of the training set with labels, forming the labeled group. The other subset contains the remaining unlabeled training data, which we defined as the unlabeled group.

3.2. implementation Details

In this study, we developed and executed our proposed algorithm using the PyTorch framework on a personal computer equipped with an NVIDIA RTX 3090 GPU. The model architecture was constructed using two Unet backbone networks, which were initialized with distinct random seeds to foster diversity in the training process. For model optimization, this work employed the Adam optimizer, with a batch size of 16, a learning rate set to 1×10^{-4}, the momentum parameters β_1 and β_2 set to 0.9 and 0.999, and the numerical stability parameter epsilon (eps) set to 1×10^{-8}.

In the design of the model's loss function, we specifically set two trade-off parameters λ_1 and λ_2, with values of 1×10^{-3} and 1, to balance the importance of different loss terms. To evaluate the performance of the model, we adopted the Dice coefficient as metrics. The training was rigorously conducted over 100 epochs.

Table 1. Comparison of the proposed model with the baseline models and state-of-the-art models.

Methods	1/2(400)	1/4(200)	1/8(100)	1/16(50)
Unet	78.56	75.63	65.87	61.24
CPC [7]	78.90	76.85	66.92	62.15
CPS [8]	79.14	75.82	66.95	62.27
PSMT [14]	81.54	80.15	68.21	64.07
ST++ [15]	81.19	79.45	71.72	66.58
Ours	**83.30**	**81.30**	**76.64**	**72.13**

3.3. Comparative Experiment

The CPC [7] and CPS [8] models are used as baselines for evaluation purposes. During the assessment phase, the same backbone network is used for training and testing both the baseline models and the models we designed. Additionally, Table 1 also shows a comparison of our model's Dice coefficient with two other state-of-the-art networks.

As shown in Table 1, among all the dataset partitions, our algorithm provided the best Dice coefficient. Compared to the baselines CPC and CPS, our model achieved nearly a 10% improvement at the 1/16 partition and a nearly 5% improvement at the 1/2 partition.

3.4. Ablation Experiment

In this study, we conducted ablation experiments to investigate the impact of fused cross semi supervised learning and gradient penalty strategies on model performance. The results, as demonstrated in Table 2, indicate that the introduction of our fusion semi-supervised learning strategy and gradient penalty mechanism significantly enhanced the model's Dice coefficient. This observation confirms the rationality of all modules in this research and substantiates that the fusion semi-supervised module indeed augments the efficacy of the gradient penalty module.

Table 2. Ablation study evaluation in Dice values of different added components, conducted on 1/8 partition.

Methods	L_{sup}	L_{grad}	L_{semi}	Dice
I	✓			65.87
II	✓	✓		73.05
III	✓		✓	66.71
Ours	✓	✓	✓	76.64

4. CONCLUSION

This study successfully developed an innovative semi supervised neural network model, GP-CPS, specifically designed for medical image segmentation tasks. The model ingeniously integrates three aspects: supervised learning, gradient penalty , and fused cross pseudo labeling. Notably, we introduce the concept of gradient penalty to the field of image segmentation for the first time, and through mathematical proof and experimental validation, fully demonstrate the significant potential of this strategy in enhancing image segmentation performance. In addition, we introduced a fused cross pseudo labeling mechanism, which not only further enhances the effect of gradient penalty but also significantly improves the quality and accuracy of pseudo labels. Through the synergistic effect of these multi-strategies, the GP-CPS model has shown outstanding performance in medical image segmentation tasks, providing new ideas and methods for future research and applications.

5. COMPLIANCE WITH ETHICAL STANDARDS

In this study, we utilized the Kvasir-SEG [13] dataset , which is an open-access collection of human subject data published for the 2020 MediaEval competition. As the data is provided with an open-access license, ethical approval is not required due to the nature of the license agreement associated with open-access data .

6. ACKNOWLEDGMENTS

No funding was received for conducting this study. The authors have no relevant financial or non-financial interests to disclose.

7. REFERENCES

[1] Shijie Hao, Yuan Zhou, and Yanrong Guo, "A brief survey on semantic segmentation with deep learning," *Neurocomputing*, vol. 406, pp. 302–321, 2020.

[2] Dong-Hyun Lee et al., "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning*, ICML. Atlanta, 2013, vol. 3, p. 896.

[3] Eric Arazo, Diego Ortego, Paul Albert, Noel E O'Connor, and Kevin McGuinness, "Pseudo-labeling and confirmation bias in deep semi-supervised learning," in *2020 International joint conference on neural networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[4] Antti Tarvainen and Harri Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *Advances in neural information processing systems*, vol.30, 2017.

[5] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii, "Virtual adversarial training: a regularization method for supervised and semi-supervised learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1979–1993, 2018.

[6] Qizhe Xie, Zihang Dai, Eduard Hovy, Thang Luong, and Quoc Le, "Unsupervised data augmentation for consistency training," *Advances in neural information processing systems*, vol. 33, pp. 6256–6268, 2020.

[7] Zhanghan Ke, Di Qiu, Kaican Li, Qiong Yan, and Rynson WH Lau, "Guided collaborative training for pixelwise semi-supervised learning," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*.Springer, 2020, pp. 429–445.

[8] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2613–2622.

[9] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6023–6032.

[10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[11] Martin Arjovsky, Soumith Chintala, and Léon Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.

[12] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville, "Improved training of wasserstein gans," *Advances in neural information processing systems*, vol. 30, 2017.

[13] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas De Lange, Dag Johansen, and Håvard D Johansen, "Kvasir-seg: A segmented polyp dataset," in *MultiMedia modeling: 26th international conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, proceedings, part II 26*. Springer, 2020, pp. 451–462.

[14] Yuyuan Liu, Yu Tian, Yuanhong Chen, Fengbei Liu, Vasileios Belagiannis, and Gustavo Carneiro, "Perturbed and strict mean teachers for semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 4258–4267.

[15] Lihe Yang, Wei Zhuo, Lei Qi, Yinghuan Shi, and Yang Gao, "St++: Make self-training work better for semi-supervised semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 4268–4277.