A Medical Image Classification Network Based on Multi-View Consistent Momentum Contrastive Learning

Chuangui Cao[1] , **Shifei Ding**[1,2**] , **Lili Guo**[1,2]
[1]School of Computer Science and Technology, China University of Mining and Technology
[2]Mine Digitization Engineering Research Center of Ministry of Education, China University of Mining and Technology
{caocg, dingsf, liliguo }@cumt.edu.cn,

Abstract

Due to variations in imaging conditions, images often exhibit discrepancies in color reproduction. Furthermore, motion-induced blur can lead to edge degradation, making color sensitivity and edge blurriness two prevalent and challenging issues in both natural image processing and medical image analysis. To address these challenges, we propose a model termed the Three-View Consistency Momentum Contrastive with Sobel Operator (SVCMC). Specifically, we first design a three-view momentum-update architecture that employs a Sobel-augmented ResNet as the backbone. We then introduce a novel contrastive loss, referred to as the Three-View Consistency Momentum Contrastive Loss. Next, to mitigate the oscillations and slow convergence commonly observed in contrastive learning, we construct a dynamic contrastive loss function that adapts in real time over the training process. Finally, we validated the superiority of our model on two medical image datasets and one natural image dataset, where its classification accuracy and convergence speed significantly outperformed existing state-of-the-art contrastive models.

1 Introduction

In natural image processing and medical image analysis, color sensitivity is a common and challenging issue. For instance, in tasks such as image classification [Krizhevsky *et al.*, 2012], object detection [Girshick *et al.*, 2014], and medical image analysis [Shen *et al.*, 2017], variations in color can significantly impact model performance, especially under conditions of lighting changes, color distortions, or the presence of noise. Additionally, in medical image analysis, color variations are often less critical compared to texture and shape, with the structure and boundaries of lesions serving as key diagnostic indicators [Litjens *et al.*, 2017]. Moreover, medical imaging data may originate from different devices, leading to color distortions and inter-device discrepancies, which makes color robustness particularly important. To address this, researchers have proposed various methods to enhance model robustness against color variations, such as color histograms [Rabie *et al.*, 2024], color jittering [Cubuk *et al.*, 2019], and Color-Aware Convolutions in deep learning [Yao *et al.*, 2025]. However, these methods often result in the loss of some image details or introduce noise due to inaccurate color representation, while also increasing computational overhead.

Another common issue is edge blur, where the details of object boundaries in images become unclear or distorted, making the visual quality of the image appear blurry and less sharp [Simonyan, 2014]. Edge information is crucial for accurately understanding and processing images. This is particularly true in tasks like image classification, object detection, and medical image analysis [Tianhao *et al.*, 2024], where edge information typically contains essential details about object shapes, structures, and spatial relationships. Consequently, edge blur can lead to errors in the model's understanding of the image, thereby affecting task accuracy. Addressing edge blur is therefore critical for enhancing model performance. Current solutions include deep deblurring networks (DeepDeblur) [Kupyn *et al.*, 2018], CycleGAN [Zhu *et al.*, 2017], and super-resolution reconstruction [Dong *et al.*, 2014]. Although these methods are effective, they generally require substantial computational resources and labeled data, and there is limited research on their application in self-supervised learning.

To tackle the issues of color sensitivity and edge blur in images, this paper proposes a contrastive model with an adaptive contrastive loss under three-view consistency constraints, incorporating the Sobel operator. This model demonstrates significant advantages within a self-supervised learning framework. Firstly, the Sobel operator extracts edge information from images and integrates it into the deep network, enhancing the representation of image edges, improving the model's sensitivity to detailed features, and increasing color robustness. Compared to existing methods, our approach not only addresses color robustness but also mitigates edge blur. Secondly, we introduce a three-view consistency momentum contrastive loss function, which enhances the consistency of representations across different views through consistency constraints, thereby improving the model's robustness to images from various angles. Momentum contrast effectively utilizes non-synchronous updated data through its

[*] Corresponding Author.

unique dynamic dictionary and encoder update strategy, enhancing the learning capacity on large-scale datasets. Considering the common challenges in contrastive learning, such as loss oscillation or slow convergence, we propose a Dynamic Contrastive Loss. This loss function dynamically adjusts with the training cycles, effectively reducing oscillations during training and accelerating model convergence. The contributions of this paper are summarized as follows:

- .We propose a three-view consistency momentum contrastive loss function that addresses color sensitivity, enhancing the model's robustness and generalization capabilities.
- We integrate the Sobel operator into the ResNet main model framework to resolve edge blur issues.
- We construct a three-view framework structure and employ momentum contrast to update view features, enriching feature extraction while enhancing stability.
- We develop a Dynamic Contrastive Loss for the training process, accelerating convergence speed and mitigating model oscillations.

2 Related Works

Edge blur and color sensitivity are key challenges in image processing and medical image analysis, significantly affecting model performance. Recent advances in edge extraction, color robustness, and loss function optimization have greatly improved model accuracy and stability. Techniques like dynamic loss, consistency loss, and momentum contrast have further enhanced training efficiency and generalization. This review highlights research on these challenges and methods.

In recent years, significant advancements have been made in the fields of image processing and medical imaging to effectively address the challenges of edge blur. These improvements have been achieved by enhancing edge extraction capabilities and ensuring clustering consistency, thus improving task performance. For example, CDANet [Yang *et al.*, 2024] adopts a dual-branch architecture that focuses on both regional and edge features, effectively enhancing edge extraction in building semantic segmentation. In the field of cancer imaging, a deep learning-based edge detection algorithm for cancer images [Li *et al.*, 2020] has successfully achieved high-precision edge detection through three-dimensional reconstruction and fine-grained feature segmentation. To further improve the accuracy of edge information, researchers have used tumor edge data as pseudo-labels for the fine-grained BI-RADS classification of breast ultrasound images [Xu *et al.*, 2024]. These studies demonstrate the critical role of edge enhancement and self-supervised learning strategies in addressing edge blur issues, particularly in enhancing the precision and robustness of image processing.

To address the challenges of color sensitivity, various innovative solutions have also been proposed in the fields of color processing and normalization in recent years. For example, the n-color balancing method [Akazawa *et al.*, 2021] was introduced to correct all colors. In the realm of color naming computation, the ColorMLP approach [Yan *et al.*, 2022] combines the RGB color model with Graph Attention Networks (GATs), learning universal color mappings and employing Partial Color Jittering (PCJ) data augmentation. In the analysis of histopathological images, the Color-Adaptive Generative Adversarial Network (CAGAN) [Cong *et al.*, 2022] was proposed for stain normalization. These methods exhibit strong application potential when dealing with complex data and diverse tasks, particularly in the areas of color correction and normalization, effectively mitigating challenges related to color sensitivity.

Dynamic loss functions are widely used in deep learning to adaptively adjust loss values during training, accelerating convergence and reducing oscillations. Early in training, larger learning rates and smaller loss weights expedite learning, while gradually lowering the learning rate later stabilizes training and avoids local minima or overfitting. For instance, Dynamic Loss Threshold (DLT) [Yang *et al.*, 2023] discards potentially incorrect labels by comparing sample losses with dynamic thresholds, greatly improving performance on noisy-label datasets. FarSeg++ [Zheng *et al.*, 2020] progressively focuses on hard samples while reducing easy-sample gradients, balancing foreground and background segmentation. Dynamic loss also enhances backdoor attacks in image compression [Yu *et al.*, 2023b] by adaptively balancing loss terms, illustrating its broad applicability in handling noisy labels, balancing sample difficulty, and bolstering robustness.

In recent years, numerous scholars have explored consistency methods[Jiang *et al.*, 2023], leading to the emergence of many innovative approaches. For instance, a deformable registration method based on paired cyclic consistent neural representations [Van Harten *et al.*, 2023] has improved accuracy and provided reliable uncertainty measurements. In semi-supervised medical image segmentation, ASE-Net [Lei *et al.*, 2022] utilizes dynamic convolution and consistency training to better align labeled and unlabeled data, thus enhancing prediction quality and reducing overfitting. Additionally, the Fuzzy Consensus Mean Teacher (AC-MT) model [Xu *et al.*, 2023] integrates fuzzy target selection to strengthen consistency learning in information-rich areas, improving segmentation outcomes. Path consistency [Lu *et al.*, 2024] enhances object matching in self-supervised learning through multiple observation paths. Together, these methods highlight the pivotal role of consistency learning in boosting model robustness and performance.

Momentum Contrast (MoCo) [He *et al.*, 2020] has shown considerable promise across various deep learning tasks, particularly in enhancing model stability, accelerating convergence, and improving generalization. The HMMC framework [Shen *et al.*, 2023] further strengthens representation generalization in text-video retrieval by integrating hierarchical matching with momentum contrast and increasing negative samples. Similarly, the USER [Zhang *et al.*, 2024] approach leverages unified semantic augmentation with momentum contrast to boost image-text retrieval. In medical imaging, momentum contrast learning combined with prototype networks and few-shot learning significantly elevates diagnostic accuracy for COVID-19 [Chen *et al.*, 2021a], while the MoMA method [Le Vuong and Kwak, 2024] employs knowledge distillation to enhance histopathological analysis. The CLEAN algorithm[Yu *et al.*, 2023a] utilizes a contrastive learning framework to assign enzyme commission (EC) numbers to enzymes.

3 Method

We first present the overall framework of the model, followed by a detailed introduction of the main innovative components within the model, including the Sobel operator, the three-view consistency momentum contrast loss function, and the Dynamic Contrastive Loss. The specific components of the model are illustrated in Figure 1.

3.1 Sobel Operator

The Sobel operator is an edge detection method that can extract the edge information of an input image while ignoring color features, thereby preventing the model from over-relying on color. This approach helps the model focus more on the structural information within the image rather than simple color features.

The Sobel operator calculates the horizontal and vertical gradients of an image using two convolution kernels, one in the horizontal direction (Gx) and the other in the vertical direction (Gy). The output of the Sobel layer is a combination of the horizontal and vertical gradients. Equations (1) and (2) respectively compute the magnitude and direction of the gradients.

$$M = \sqrt{G_x^2 + G_{\mathrm{y}}^2} \tag{1}$$

Direction: Represents the orientation of the edge and is calculated as the arctangent of the gradient.

$$\theta = a\tan 2(G_y, G_x) \tag{2}$$

In DeepCluster, Sobel filters were employed to preprocess the input in order to prevent the model from merely relying on color information for clustering, instead encouraging the use of more meaningful features such as edges and shapes. Building on this approach, we incorporate Sobel layers into the ResNet architecture to enhance its focus on structural information within pathology images. This integration helps alleviate color sensitivity issues in pathology image classification, thereby improving the model's generalization capability. Additionally, ResNetSobel supports the use of deep stem layers, which are analogous to the deep stem layers in the standard ResNet. Depending on the model configuration, the input may first pass through the stem layer or directly utilize the initial convolution, batch normalization, and ReLU activation.

3.2 Three-view Consistency Momentum Contrastive Loss

Consistency loss functions are widely applied in semi-supervised and self-supervised learning. Inspired by the infoNCE consistency loss function, we extend it to a three-view framework by employing three different data augmentations (such as cropping, rotation, and color jittering) for contrastive learning of samples. The goal is to encourage the model to maintain consistent prediction results for the same input data under different views or data transformations. This is achieved by minimizing the differences in the model's outputs for identical samples under varying conditions, thereby enhancing the model's robustness and generalization capabilities. We refer to this as the three-view consistency contrastive loss function, as detailed in Equations (3-7).

$$Lv_1, v_2 = -\log \frac{\exp(sim(f(v_1), f(v_2))/t}{\sum_{k=1}^{N} \exp(sim(f(v_1), f(x_k))/t} \tag{3}$$

$$Lv_2, v_3 = -\log \frac{\exp(sim(f(v_2), f(v_3))/t}{\sum_{k=1}^{N} \exp(sim(f(v_2), f(x_k))/t} \tag{4}$$

$$Lv_3, v_1 = -\log \frac{\exp(sim(f(v_3), f(v_1))/t}{\sum_{k=1}^{N} \exp(sim(f(v_3), f(x_k))/t} \tag{5}$$

$$L_{\text{total}} = Lv_1, v_2 + Lv_2, v_3 + Lv_3, v_1 \tag{6}$$

In which t is a temperature hyperparameter used to control the scale of similarity.

We construct three distinct views$f(v_1)$, $f(v_2)$, $f(v_3)$ and perform cross-contrastive comparisons among them. Specifically, Lv_1, v_2 measures the similarity between view1 and view2, while Lv_2, v_3 and Lv_3, v_1 are computed in a similar manner. The contrastive loss encourages these similarities to be as close as possible, thereby achieving feature alignment by minimizing the distances between them. Finally, the three loss components are weighted and summed to obtain the final three-view consistency contrastive loss function.

We introduce a momentum contrast update mechanism to update the second views ($f(v_1)$, $f(v_2)$, $f(v_3)$), utilizing momentum updates to smooth the updating process and prevent drastic fluctuations in model parameters, as shown in Equation (7).

$$f(v_{i\_mc}) = \beta f(v_i) + (1 - \beta)g \tag{7}$$

Specifically, $i = 1, 2, 3$, the default value β = 0.99.

In self-supervised learning, a momentum encoder is a technique used to enhance training stability and performance by maintaining smooth updates of the model's historical parameters. A higher momentum coefficient indicates slower updates, resulting in greater retention of historical information. The default value is set to β = 0.99, which means that the encoder parameters are updated very minimally, allowing historical information to dominate. Momentum updates enhance stability during the optimization process by leveraging information from previous updates to adjust the current parameters. This enables the parameters to be "accelerated" along the gradient direction to some extent, thereby facilitating faster model convergence. The three-view consistency contrastive loss function, integrated with momentum updates, can be expressed as Equation (8):

$$Lv_i, v_{j\_mc} = -\log \frac{\exp(sim(f(v_i), f(v_{j\_mc}))/t}{\sum_{k=1}^{N} \exp(sim(f(v_i), f(x_k))/t} \tag{8}$$

Specifically, $i = 1, 2, 3; j = 2, 3, 1$

Summing these terms results in the three-view consistency momentum contrast loss function, as shown in Equation (9).

$$L_{\text{total\_mc}} = Lv_1, v_{2\_mc} + Lv_2, v_{3\_mc} + Lv_3, v_{1\_mc} \tag{9}$$

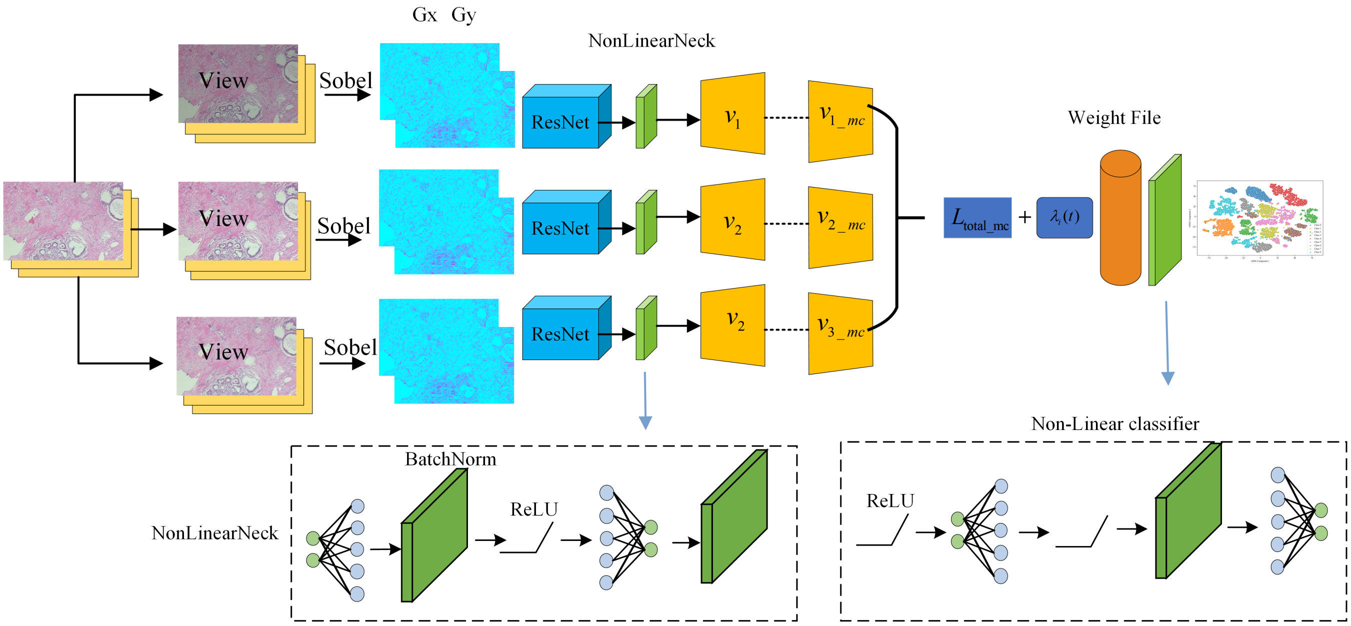


Figure 1: The proposed SVCMC learning framework

3.3 Dynamic Contrastive Loss

To accelerate model convergence and mitigate training oscillations, we have developed a Dynamic Contrastive Loss. This loss function dynamically adjusts the loss between each pair of views based on the current training stage or model performance. This approach enables the model to adaptively adjust the weights of different view pairs during the training process, thereby allowing the training to flexibly and efficiently focus on the most important view information.

Dynamic contrastive loss refers to the dynamic adjustment of the loss function's weights or other parameters based on the training progress or current feedback from the model. We adjust the loss function weights corresponding to each view pair according to their current training performance. With the incorporation of dynamic loss, our total loss function is expressed in Equation (10):

$$\begin{aligned} L^{d}{}_{\text{total\_mc}} = \lambda_1(t)Lv_1, v_{2\_mc} + \lambda_2(t)Lv_2, v_{3\_mc} \\ + \lambda_3(t)Lv_3, v_{1\_mc} \end{aligned} \tag{10}$$

Specifically, $\lambda_1(t), \lambda_2(t), \lambda_3(t)$ is a dynamic weight function that varies over time (training steps).

If the loss value of a particular view pair is large, it indicates that the model's learning progress for this view is slow, and it may be necessary to dynamically decrease its loss weight. The specific adjustment method is detailed in Equation (11).

$$\lambda_i(t) = \frac{1}{1 + \alpha_i Lv_i, v_{j\_mc}(t)} \tag{11}$$

Where, $i = 1, 2, 3$, $j = 2, 3, 1$

The hyperparameter $\alpha = 0.1$. As shown in Equation (11), the variation of $\lambda_i(t)$ is determined by the loss. When the loss value is high, indicating a large error concerning the true value, we dynamically adjust it to reduce the loss.

Similarly, after incorporating it into the three-view consistency contrastive loss function, as shown in Equation (12):

$$L^{d}{}_{\text{total}} = \lambda_1(t)Lv_1, v_2 + \lambda_2(t)Lv_2, v_3 + \lambda_3(t)Lv_3, v_1 \tag{12}$$

In the subsequent experimental section, we will separately employ the two different loss functions, $L^{d}{}_{\text{total}}$ and $L^{d}{}_{\text{total\_mc}}$, to conduct experimental validations.

4 Experiments

4.1 Datasets

To effectively validate the generalization and robustness of our model, we conducted comprehensive evaluations using three publicly available pathology image datasets: the multiscale BreaKHis dataset [Spanhol *et al.*, 2015] and the large-scale pathology dataset NCT-CRC-HE-100K [Kather *et al.*, 2018]. Additionally, we utilized the natural image dataset Food-101 [Bossard *et al.*, 2014] to assess the model's performance across different domains and tasks. By employing these diverse datasets, we are able to thoroughly evaluate the model's adaptability and stability, ensuring its wide applicability in real-world applications.

4.2 Parameter setting

To ensure a more accurate and unbiased evaluation of the model's performance, we standardized several parameter settings, as shown in Table 1.

Parameter	Value
Momentum Coefficient	0.99
Epoch	200
Optimizer	LARS
Batch Size	32/64
Learning Rate Decay	Cosine Annealing

Table 1: Parameter setting

We used data augmentation techniques including RandomResizedCrop, ColorJitter, and RandomFlip. RandomResizedCrop first performs random cropping on the image and then resizes it to the target dimensions of 224×224. This method effectively increases data diversity, thereby enhancing the robustness of the model. ColorJitter randomly adjusts the image's brightness, contrast, hue, and saturation, generating images with varied visual effects. This helps the model adapt to different lighting and color conditions, thereby improving its generalizability. RandomFlip applies horizontal or vertical flips to images with a certain probability, assisting the model in handling scenarios where flipping does not affect classification outcomes, thereby further enhancing its robustness. Through these data augmentation methods, the model is better able to adapt to complex environments, improving performance and stability during the training process.

4.3 Comparative Experiments

Firstly, we conducted comparative experiments on the BreaKHis pathology image dataset at various magnification scales. The objective was to evaluate the model's performance across different magnification levels of the pathology dataset and to investigate the model's color robustness and edge detection capabilities. The detailed experimental results are shown in Table 2.

Method	ACC	Pre	F1
MoCo v1[He *et al.*, 2020]	97.98	98.07	97.98
MoCo v2[Chen *et al.*, 2020b]	87.80	88.74	87.60
MoCo v3[Chen *et al.*, 2021b]	89.32	90.47	89.37
SimCLR[Chen *et al.*, 2020a]	97.47	97.74	97.52
SimSiam[Chen and He, 2021]	98.61	98.68	98.60
MPCS[Chhipa *et al.*, 2023]	92.18	-	-
Breast-NET[Saha *et al.*, 2024]	90.34	91.00	90.00
SVCMC	**99.37**	**99.40**	**99.37**

Table 2: Comparative experiments on the BreaKHis dataset

While self-supervised methods like MoCo and SimCLR excel in breast cancer image classification, they mainly focus on global features, potentially overlooking subtle edge and color differences. In contrast, our model integrates the Sobel operator and Dynamic Contrastive Loss, enhancing its ability to capture local features and improve generalization and representation.

The underlying reason for this performance improvement lies in the integration of the ResNet architecture with the Sobel operator and the three-view consistency enhanced momentum contrast loss function. The Sobel operator enhances the model's sensitivity to edge information, the ResNet architecture ensures the depth and diversity of feature extraction, and the momentum contrast loss function combined with multi-view consistency enhancement further improves training stability, model robustness, and the effectiveness of self-supervised learning. The combination of these innovative methods enables our model to achieve outstanding performance across multiple datasets.

Subsequently, we conducted experiments on a large pathology image dataset to explore the model's classification performance on large-scale data. The specific experimental results are presented in Table 3.

Method	ACC	Pre	F1
MoCo v1 [He *et al.*, 2020]	97.38	97.39	97.36
MoCo v2 [Chen *et al.*, 2020b]	97.14	97.52	97.18
MoCo v3 [Chen *et al.*, 2021b]	96.57	97.13	96.62
SimCLR [Chen *et al.*, 2020a]	98.44	98.46	99.45
SimSiam [Chen and He, 2021]	95.46	99.91	97.46
HistoSSL-vit [Jin *et al.*, 2022]	96.18	-	-
Contrastive [Chu *et al.*, 2023]	88.12	-	-
iDeComp [Buczek *et al.*, 2023]	94.90	95.00	95.00
SAG-ViT [Shravan *et al.*, 2024]	98.61	-	-
SVCMC	**99.98**	**99.98**	**99.98**

Table 3: Comparative experiments on the NCT-CRC-HE-100K dataset

From the perspective of capturing edge information and color sensitivity, Mymodel improves the sensitivity to image details, structure, and color variations by combining the Sobel operator and color jittering. This allows the model to not only accurately identify critical edge features on the NCT-CRC-HE-100K dataset but also adapt to varying lighting and color conditions, resulting in a 99.98% accuracy and the best performance across other metrics. In comparison, while other methods perform well on certain metrics, they do not enhance edge information and color sensitivity as effectively as our model, which leads to slightly inferior performance in handling details and complex environments.

Finally, we conducted experiments on the large-scale dataset Food-101 to explore the model's performance in natural image classification, demonstrating its robustness and generalization ability. The specific experimental results are shown in Table 4.

Method	ACC	Pre	F1
MoCo v1 [He *et al.*, 2020]	97.38	97.39	97.36
MoCo v2 [Chen *et al.*, 2020b]	97.14	97.52	97.18
MoCo v3 [Chen *et al.*, 2021b]	96.57	97.13	96.62
SimCLR [Chen *et al.*, 2020a]	98.44	98.46	99.45
SimSiam [Chen and He, 2021]	95.46	99.91	97.46
HistoSSL-vit [Jin *et al.*, 2022]	96.18	-	-
Contrastive [Chu *et al.*, 2023]	88.12	-	-
iDeComp [Buczek *et al.*, 2023]	94.90	95.00	95.00
SAG-ViT [Shravan *et al.*, 2024]	98.61	-	-
SVCMC	**99.98**	**99.98**	**99.98**

Table 4: Comparative experiments on the Food-101 dataset

Our model not only outperforms other models in terms of accuracy but also demonstrates equally outstanding performance across other key metrics, including recall, precision, and F1 score. This indicates that the model is highly effective in distinguishing between different categories, with extremely high sensitivity and specificity, resulting in very few false positives and false negatives. At the same time, SimSiam and other more recent models have also shown competitive performance.

The experimental results demonstrate that our model successfully addresses the common issues of edge blurring and color sensitivity in medical image analysis through hierarchical feature enhancement, dynamic optimization mechanisms, and the Sobel operator, while maintaining excellent performance in natural scenes.

4.4 Ablation Experiments

To better validate the significant roles of the (Dynamic Contrastive Loss, DCL)and(Momentum Contrastive Loss, MCL) within the model, we designed the ablation experiments shown in Table 5.

Method	ACC	Pre	F1
SVCMC w/o DCL	97.41	97.56	97.31
SVCMC w/o MCL	97.28	97.71	97.34
SVCMC w/o DCL and MCL	96.96	96.87	96.93
SVCMC	**99.37**	**99.40**	**99.37**

Table 5: Ablation Experiments on the BreaKHis dataset

Based on the results of the ablation study, we can draw the following analysis:

After removing the dynamic loss module, all metrics slightly decline, with accuracy dropping by 1.96 percentage points. The equal distribution of loss weights across all samples leads to insufficient optimization of challenging samples, such as those with blurred edges, reducing overall detection accuracy.

When the momentum contrastive module is removed, the model performance drops slightly, with accuracy of 97.28%, a decrease of 2.09 percentage points. Compared to the removal of the dynamic loss module, the effect of removing the momentum update module is more pronounced, particularly with an improvement in precision (Pre) and F1 score. However, this improvement may come at the cost of stability.

When the SobelResNet module is removed, the model performance drops significantly, with accuracy of 97.16%, a decrease of 2.21 percentage points. The SobelResNet module plays a crucial role in edge feature extraction in images, and the lack of suppression of color sensitivity leads to a noticeable decline in performance.

4.5 Discussion on Backbone and Classifier

To further discuss the role of the backbone and classifier in classification, we conducted experiments on three datasets and provided a detailed discussion. The specific experimental results are shown in Table 6 and Table 7.

The classification results show that SobelResNet outperforms ResNet, mainly due to the strong reliance of pathological images on morphological structures, where staining inconsistencies can affect performance. SobelResNet reduces the dependency on color and emphasizes texture and structural features. Similarly, in natural images, factors like lighting variations can introduce noise. SobelResNet's focus on texture and structural features helps mitigate the effects of color variations, making it more robust and adaptable for both pathological and natural image classification tasks.

Backbone	Dataset	ACC	Pre	F1
SobelResNet	BreaKHis	99.37	99.40	99.37
ResNet	BreaKHis	95.89	96.43	96.16
SobelResNet	NCT100K	99.98	99.98	99.98
ResNet	NCT100K	98.70	98.72	98.71
SobelResNet	food-101	99.38	99.39	99.37
ResNet	food-101	95.71	95.71	95.93

Table 6: Discussion on Backbone, NCT-CRC-HE-100K(NCT100k)

Classifier	Dataset	ACC	Pre	F1
Linear	BreaKHis	96.21	96.55	96.19
Nonlinear	BreaKHis	99.37	99.40	99.37
Linear	NCT100K	99.40	99.39	99.38
Nonlinear	NCT100K	99.98	99.98	99.98
Linear	food-101	99.36	99.34	99.35
Nonlinear	food-101	99.38	99.39	99.37

Table 7: Discussion on Classifier, NCT-CRC-HE-100K(NCT100k)

For larger datasets, the difference between linear and nonlinear models is relatively minor, and we consider the variation to be within an acceptable error margin. However, the significant disparity observed in the BreaKHis dataset is likely due to the presence of images captured at different magnifications. Nonlinear models can, to some extent, mitigate misclassifications caused by variations in magnification, providing better robustness under such conditions.

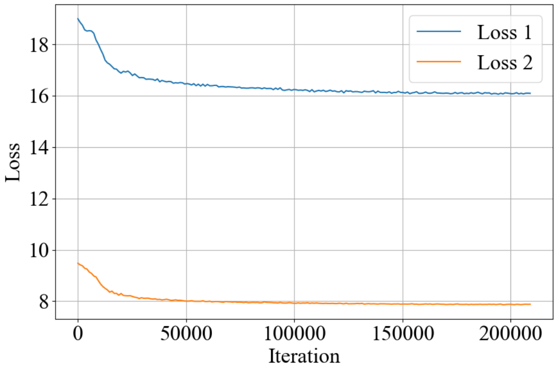
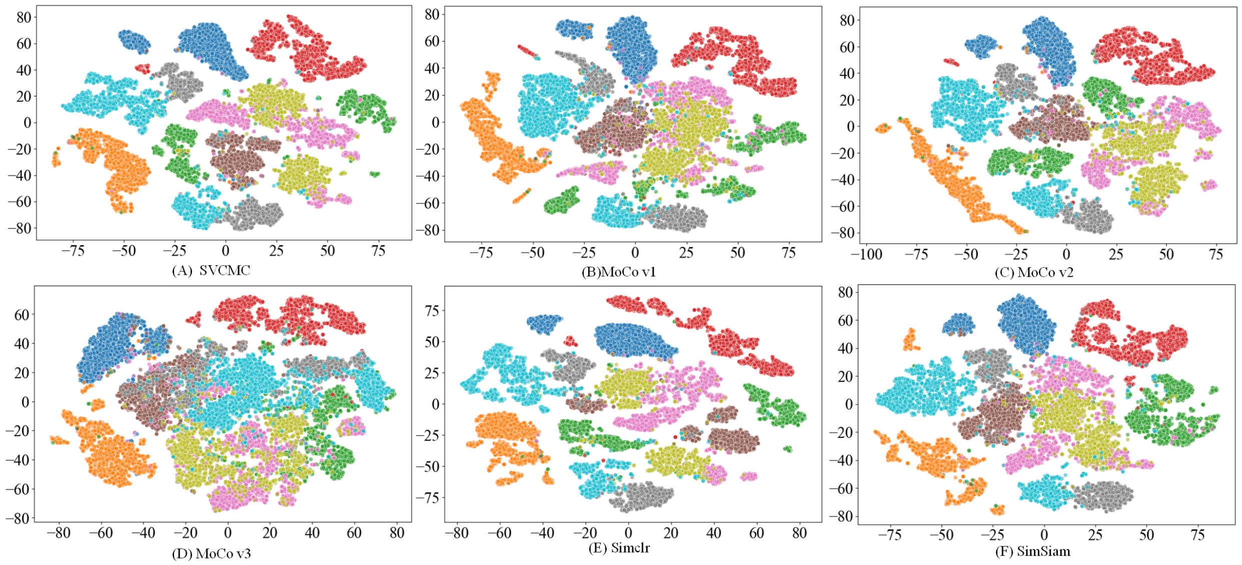


Figure 2: loss curve in the graph

4.6 Visualization

In Figure 2, loss2, which incorporates Dynamic Contrastive Loss, exhibits a faster convergence rate: In the initial stages

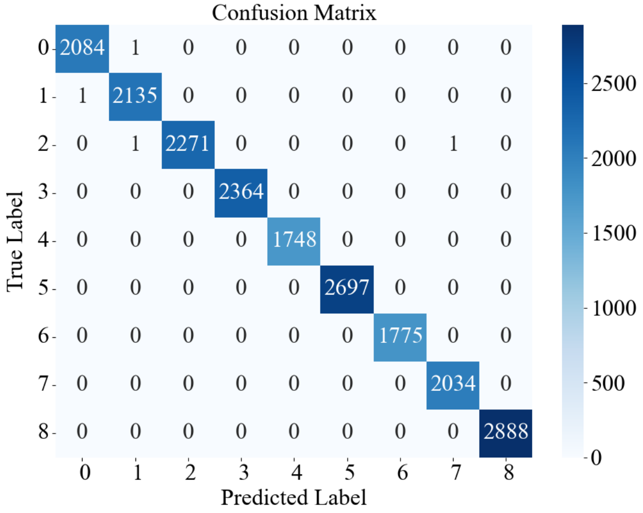


(A) SVCMC (B)MoCo v1 (C) MoCo v2

(D) MoCo v3 (E) Simclr (F) SimSiam

Figure 3: T-SNE Visualization

of iteration, loss2 demonstrates a quicker decline in loss compared to loss1, which is highly beneficial for saving training time and resources. Higher stability: Throughout the training process, loss2 shows greater stability, which helps to prevent performance degradation due to overfitting, particularly important in training with large-scale data sets. Lower final loss values: The lower final loss values indicate that under the same training conditions, loss2 may optimize model parameters more effectively.



	0	1	2	3	4	5	6	7	8
0	2084	1	0	0	0	0	0	0	0
1	1	2135	0	0	0	0	0	0	0
2	0	1	2271	0	0	0	0	1	0
3	0	0	0	2364	0	0	0	0	0
4	0	0	0	0	1748	0	0	0	0
5	0	0	0	0	0	2697	0	0	0
6	0	0	0	0	0	0	1775	0	0
7	0	0	0	0	0	0	0	2034	0
8	0	0	0	0	0	0	0	0	2888

Figure 4: Confusion Matrix

As shown in Figure 3, this study employs the T-SNE visualization method to analyze the classification performance across 10 classes (Class 0–9), each represented by a different color. Figure 3(A) exhibits clear inter-class separability, where the data points of each class form distinct and well-defined clusters. This indicates that data points from different classes are likely to be highly distinguishable in the original high-dimensional space, providing strong evidence of the model's discriminative power and classification capability in the high-dimensional feature space. In panel (D), most of the classes exhibit compact clusters, suggesting that data points within the same class are highly similar in the original feature space. A few classes, such as the brown and orange clusters, show some overlap or proximity, which may indicate that these classes share certain similarities in the feature space.

As shown in the confusion matrix in Figure 4, most of the predictions are concentrated along the diagonal, exhibiting a clear pattern of centralized distribution. The values on the main diagonal are significantly higher than those in the off-diagonal regions, intuitively reflecting the model's high classification accuracy across the majority of classes. Although a few instances of inter-class misclassification are observed, their overall number is limited, further indicating that the model achieves good classification performance on this dataset.

5 Conclusion

The SVCMC algorithm developed in this study incorporates the Sobel operator and introduces a novel three-view consistency momentum contrast loss function, along with Dynamic Contrastive Loss. This approach has been rigorously tested across two pathological datasets and one natural image dataset, where it has demonstrated significant strengths in classification performance. The model excels in effectively managing issues related to color sensitivity and edge blur in images, showcasing its robustness and adaptability. Looking ahead, plans are in place to extend this research to include more comprehensive experiments on remote sensing imagery and a variety of structured medical datasets, aiming to further validate and enhance the algorithm's applicability and performance across diverse imaging contexts.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant no.62276265 and 62406326).

References

[Akazawa *et al.*, 2021] Teruaki Akazawa, Yuma Kinoshita, and Hitoshi Kiya. Multi-color balance for color constancy. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 1369–1373. IEEE, 2021.

[Bossard *et al.*, 2014] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101–mining discriminative components with random forests. In *Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part VI 13*, pages 446–461. Springer, 2014.

[Buczek *et al.*, 2023] Patryk Buczek, Usama Zidan, Mohamed Medhat Gaber, and Mohammed M Abdelsamea. Idecomp: imbalance-aware decomposition for class-decomposed classification using conditional gans. *Discover Artificial Intelligence*, 3(1):31, 2023.

[Chen and He, 2021] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15750–15758, 2021.

[Chen *et al.*, 2020a] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[Chen *et al.*, 2020b] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020.

[Chen *et al.*, 2021a] Xiaocong Chen, Lina Yao, Tao Zhou, Jinming Dong, and Yu Zhang. Momentum contrastive learning for few-shot covid-19 diagnosis from chest ct images. *Pattern recognition*, 113:107826, 2021.

[Chen *et al.*, 2021b] Xinlei Chen, Saining Xie, and Kaiming He. An empirical study of training self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9640–9649, 2021.

[Chhipa *et al.*, 2023] Prakash Chandra Chhipa, Richa Upadhyay, Gustav Grund Pihlgren, Rajkumar Saini, Seiichi Uchida, and Marcus Liwicki. Magnification prior: a self-supervised method for learning representations on breast cancer histopathological images. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2717–2727, 2023.

[Chu *et al.*, 2023] Hongbo Chu, Fang Li, Yonghong He, and Tian Guan. Generative and contrastive based self-supervised learning model for histopathology image analysis. In *Proceedings of the 2023 15th International Conference on Machine Learning and Computing*, pages 354–360, 2023.

[Cong *et al.*, 2022] Cong Cong, Sidong Liu, Antonio Di Ieva, Maurice Pagnucco, Shlomo Berkovsky, and Yang Song. Colour adaptive generative networks for stain normalisation of histopathology images. *Medical Image Analysis*, 82:102580, 2022.

[Cubuk *et al.*, 2019] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 113–123, 2019.

[Dong *et al.*, 2014] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland*, pages 184–199. Springer, 2014.

[Girshick *et al.*, 2014] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[He *et al.*, 2020] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[Jiang *et al.*, 2023] Shenwang Jiang, Jianan Li, Jizhou Zhang, Ying Wang, and Tingfa Xu. Dynamic loss for robust learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12):14420–14434, 2023.

[Jin *et al.*, 2022] Xu Jin, Teng Huang, Ke Wen, Mengxian Chi, and Hong An. Histossl: Self-supervised representation learning for classifying histopathology images. *Mathematics*, 11(1):110, 2022.

[Kather *et al.*, 2018] Jakob Nikolas Kather, Niels Halama, and Alexander Marx. 100,000 histological images of human colorectal cancer and healthy tissue, May 2018.

[Krizhevsky *et al.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[Kupyn *et al.*, 2018] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018.

[Le Vuong and Kwak, 2024] Trinh Thi Le Vuong and Jin Tae Kwak. Moma: momentum contrastive learning with multi-head attention-based knowledge distillation for histopathology image analysis. *Medical Image Analysis*, page 103421, 2024.

[Lei *et al.*, 2022] Tao Lei, Dong Zhang, Xiaogang Du, Xuan Wang, Yong Wan, and Asoke K Nandi. Semi-supervised medical image segmentation using adversarial consistency learning and dynamic convolution network. *IEEE transactions on medical imaging*, 42(5):1265–1277, 2022.

[Li *et al.*, 2020] Xiaofeng Li, Hongshuang Jiao, and Yanwei Wang. Edge detection algorithm of cancer image based on deep learning. *Bioengineered*, 11(1):693–707, 2020.

[Litjens *et al.*, 2017] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[Lu *et al.*, 2024] Zijia Lu, Bing Shuai, Yanbei Chen, Zhenlin Xu, and Davide Modolo. Self-supervised multi-object tracking with path consistency. 2024.

[Rabie *et al.*, 2024] Tamer Rabie, Mohammed Baziyad, Radhwan Sani, Talal Bonny, and Raouf Fareh. Color histogram contouring: a new training-less approach to object detection. *Electronics*, 13(13):2522, 2024.

[Saha *et al.*, 2024] Mousumi Saha, Mainak Chakraborty, Suchismita Maiti, and Deepanwita Das. Breast-net: a lightweight dcnn model for breast cancer detection and grading using histological samples. *Neural Computing and Applications*, 36(32):20067–20087, 2024.

[Shen *et al.*, 2017] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19(1):221–248, 2017.

[Shen *et al.*, 2023] Wenxue Shen, Jingkuan Song, Xiaosu Zhu, Gongfu Li, and Heng Tao Shen. End-to-end pretraining with hierarchical matching and momentum contrast for text-video retrieval. *IEEE Transactions on Image Processing*, 32:5017–5030, 2023.

[Shravan *et al.*, 2024] Venkatraman Shravan, Jaskaran Singh Walia, et al. Sag-vit: A scale-aware, high-fidelity patching approach with graph attention for vision transformers. *arXiv preprint arXiv:2411.09420*, 2024.

[Simonyan, 2014] Karen Simonyan. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[Spanhol *et al.*, 2015] Fabio A Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast cancer histopathological image classification. *Ieee transactions on biomedical engineering*, 63(7):1455–1462, 2015.

[Tianhao *et al.*, 2024] Bu Tianhao, Michalis Lazarou, and Tania Stathaki. Image edge enhancement for effective image classification. *Proceedings Copyright*, 444:451, 2024.

[Van Harten *et al.*, 2023] Louis D Van Harten, Jaap Stoker, and Ivana Išgum. Robust deformable image registration using cycle-consistent implicit representations. *IEEE Transactions on Medical Imaging*, 2023.

[Xu *et al.*, 2023] Zhe Xu, Yixin Wang, Donghuan Lu, Xiangde Luo, Jiangpeng Yan, Yefeng Zheng, and Raymond Kai-yu Tong. Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation. *Medical Image Analysis*, 88:102880, 2023.

[Xu *et al.*, 2024] Meng Xu, Jianhua Huang, Kuan Huang, and Feifei Liu. Incorporating tumor edge information for fine-grained bi-rads classification of breast ultrasound images. *IEEE Access*, 2024.

[Yan *et al.*, 2022] Zipei Yan, Linchuan Xu, Atsushi Suzuki, Jing Wang, Jiannong Cao, and Jun Huang. Rgb color model aware computational color naming and its application to data augmentation. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 1172–1181. IEEE, 2022.

[Yang *et al.*, 2023] Hao Yang, You-Zhi Jin, Zi-Yin Li, Deng-Bao Wang, Xin Geng, and Min-Ling Zhang. Learning from noisy labels via dynamic loss thresholding. *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[Yang *et al.*, 2024] Mengyuan Yang, Rui Yang, Shikang Tao, Xin Zhang, and Min Wang. Unsupervised domain adaptive building semantic segmentation network by edge-enhanced contrastive learning. *Neural Networks*, 179:106581, 2024.

[Yao *et al.*, 2025] Jiaxin Yao, Yongqiang Zhao, Yuanyang Bu, Seong G Kong, and Xun Zhang. Color-aware fusion of nighttime infrared and visible images. *Engineering Applications of Artificial Intelligence*, 139:109521, 2025.

[Yu *et al.*, 2023a] Tianhao Yu, Haiyang Cui, Jianan Canal Li, Yunan Luo, Guangde Jiang, and Huimin Zhao. Enzyme function prediction using contrastive learning. *Science*, 379(6639):1358–1363, 2023.

[Yu *et al.*, 2023b] Yi Yu, Yufei Wang, Wenhan Yang, Shijian Lu, Yap-Peng Tan, and Alex C Kot. Backdoor attacks against deep image compression via adaptive frequency trigger. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12250–12259, 2023.

[Zhang *et al.*, 2024] Yan Zhang, Zhong Ji, Di Wang, Yanwei Pang, and Xuelong Li. User: Unified semantic enhancement with momentum contrast for image-text retrieval. *IEEE Transactions on Image Processing*, 2024.

[Zheng *et al.*, 2020] Zhuo Zheng, Yanfei Zhong, Junjue Wang, and Ailong Ma. Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4096–4105, 2020.

[Zhu *et al.*, 2017] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.