Adaptive Temperature DWA: A Flexible Task-Weighting Strategy for Simultaneous Tumor Segmentation and Classification

Muhammad Zaki Mubarak Hariyadi
Department of Robotics and Artificial Intelligence
Universitas Airlangga
Surabaya, Indonesia
zakimubarak3108@gmail.com

Yulius Harjoseputro*
Department of Informatics
Universitas Atma Jaya Yogyakarta
Yogyakarta, Indonesia
yulius.harjoseputro@uajy.ac.id

Yung-Yao Chen
Department of Electronics and Computer Engineering
National Taiwan University of Science and Technology
Taipei, Taiwan
yungyaochen@gapps.ntust.edu.tw

Aloysius Bagas Pradipta Irianto
Department of Informatics
Universitas Atma Jaya Yogyakarta
Yogyakarta, Indonesia
bagas.pradipta@uajy.ac.id

Warawut Janwittayachai
Department of Electronics and Computer Engineering
National Taiwan University of Science and Technology
Taipei, Taiwan
warawoot.jan.edu@gmail.com

***Abstract*— Multi-task learning (MTL) improves generalization and computational efficiency by learning shared representations across related tasks, yet imbalanced task learning often leads to unstable optimization. Dynamic Weight Averaging (DWA) addresses this challenge by reweighting tasks based on relative loss changes; however, its fixed temperature limits adaptability as learning dynamics evolves. To overcome this limitation, we propose Adaptive Temperature Dynamic Weight Averaging (ATDWA), a lightweight extension that adjusts the temperature using the temporal standard deviation of each task's loss, enabling more responsive and stable task weighting. We integrate ATDWA into a joint segmentation–classification framework with a pretrained ResNet-34 backbone and evaluate it on the Breast Ultrasound Images (BUSI) dataset. Compared with the original SSC, ATDWA improves segmentation performance, increasing Dice and IoU by 5.4986% and 1.4863%, respectively, while segmentation accuracy decreases slightly by 0.5138%. For classification, ATDWA further boosts accuracy, Recall, and Precision by 5.8993%, 5.8993%, and 5.4230%, respectively. Overall, these results indicate that ATDWA stabilizes multi-task optimization and substantially enhances diagnostic performance in breast ultrasound analysis.**

***Keywords*— multi-task learning, task imbalance, adaptive temperature, dynamic weight average, breast ultrasound images.**

I. INTRODUCTION

Multi-task learning (MTL) has emerged as a crucial paradigm in deep learning, enabling a single model to learn multiple related tasks simultaneously. The primary motivation for using MTL is its ability to reduce overfitting and data scarcity by leveraging shared representations across tasks, resulting in better overall generalization [1, 2]. Furthermore, MTL provides computational advantages, as a unified model typically requires fewer resources and memory than training separate models for each task [3].

However, most multi-task learning networks face a fundamental challenge: training multiple tasks simultaneously is difficult without striking an appropriate balance [4–6]. When one task dominates the optimization process, the remaining tasks may receive insufficient learning signals, resulting in decreased overall performance. Dynamic Weight Averaging (DWA) [4] was introduced to address this issue by adjusting task weights based on changes in relative loss, thereby providing a lightweight mechanism for balancing task contributions during training. However, because DWA uses a fixed temperature hyperparameter, it is unable to adapt when task behaviours change over time, limiting its responsiveness to instability in task learning dynamics.

To address this limitation, we present Adaptive Temperature Dynamic Weight Averaging (ATDWA), an extension of DWA that incorporates temporal uncertainty into the weighting mechanism. Instead of using a static temperature, ATDWA dynamically adjusts the temperature based on the recent volatility of each task's loss, measured as the standard deviation over a small epoch window. By averaging these task-specific uncertainties, ATDWA estimates the overall instability of the

learning process and proportionally increases the temperature as losses fluctuate. This adaptive strategy enables the weighting mechanism to respond more effectively to unpredictable or divergent task behaviours, resulting in more stable and context-aware weight updates throughout training.

Notably, ATDWA is architecture-agnostic because it relies only on general multi-task signals, namely task-wise losses and their relative learning dynamics, making it a potentially general-purpose task-weighting method. In this study, we evaluate ATDWA for medical image analysis by integrating it into the Simultaneous Segmentation and Classification (SSC) framework [7]. Since the original SSC encoder tends to overfit and generalize poorly, we replace it with an ImageNet-pretrained ResNet-34 (SSC-ResNet), which provides more robust and discriminative features [11, 16], mitigates vanishing gradients through residual learning [8–10], and improves convergence speed and training stability [12, 21]. ResNet-34 is also widely validated as an effective feature extractor for multi-task learning and medical image analysis [14, 15].

The primary goal of this study is to create and validate an adaptive loss-weighting strategy that improves optimization stability in simultaneous tumor segmentation and classification. The proposed method introduces Adaptive Temperature Dynamic Weight Averaging ATDWA, which dynamically regulates task importance based on temporal loss uncertainty, enabling more balanced learning. This mechanism reduces task dominance, increases convergence stability, and improves overall performance in multi-task medical imaging. The primary findings of this study are summarised as follows:

- The proposed ATDWA introduces an adaptive temperature mechanism based on the temporal variability of task losses, enabling more uncertainty-aware and responsive task weighting in multi-task learning.
- The integration of ATDWA into the SSC-ResNet framework improves both segmentation and classification performance for breast ultrasound imaging, surpassing the original SSC [7] and all SSC-ResNet variants, with and without adaptive weighting.

The remainder of this paper is structured as follows. Section II provides a detailed literature review on multi-task learning in medical imaging and task-weighting strategies. Section III presents the methodology, including details of the proposed ATDWA and SSC-ResNet framework. Section IV describes the experimental setup and reports the results. Finally, Section V concludes the paper by highlighting the main contributions, discussing limitations, and outlining directions for future research.

II. Related Works

Multi-task learning (MTL) has been widely adopted in medical image analysis for tasks such as segmentation and classification. It has been shown to enhance diagnostic robustness by jointly leveraging pixel- and image-level information across modalities, including ultrasound, mammography, dermoscopy, and CT imaging [7, 17, 18]. While MTL frameworks often outperform single-task models, jointly optimizing heterogeneous tasks remains challenging due to task imbalance, where dominant tasks suppress the learning of others, leading to unstable convergence and inconsistent performance, particularly in complex anatomical settings [4–6].

To address task imbalance, several adaptive loss-weighting strategies for multi-task learning have been proposed. These methods dynamically adjust task contributions during training to ensure that optimization is stable and that no single task dominates the shared representation. Uncertainty-based weighting [5] adjusts task importance according to homoscedastic uncertainty by scaling each task loss, thereby reducing the influence of tasks with higher predictive uncertainty during optimization. However, its effectiveness is heavily reliant on accurate uncertainty estimation, which may not always reflect actual task learning progress. GradNorm [6] aims to maintain balanced learning speeds by equating gradient magnitudes across tasks; however, this method incurs significant computational overhead due to repeated gradient manipulations. Although these methods improve static weighting schemes, they are still unable to adapt smoothly to rapid or irregular fluctuations in task difficulty while training.

Dynamic Weight Averaging (DWA) [4] offers a lightweight alternative by adjusting task weights based on relative loss changes across epochs, without requiring gradient normalization or explicit uncertainty modeling. By emphasizing tasks whose losses decrease more slowly, DWA helps mitigate the learning speed imbalance. However, its reliance on a constant temperature parameter limits adaptability, as fixed sensitivity may lead to under- or overcompensation when task dynamics vary across training stages.

To overcome this limitation, we propose Adaptive Temperature Dynamic Weight Averaging (ATDWA), which extends DWA by dynamically adjusting the temperature according to the temporal variability of task losses. When recent loss fluctuations are high, ATDWA increases the temperature to smooth weight updates and reduce instability. As training becomes more stable, the temperature decreases, enabling more responsive adjustments to meaningful differences in task learning progress.

This adaptive weighting mechanism better aligns with evolving task dynamics and maintains balanced optimization throughout training. Integrated into the SSC-ResNet framework, ATDWA consistently outperforms the original SSC, SSC-ResNet without adaptive weighting, and SSC-ResNet with existing strategies such as GradNorm and standard DWA. These results demonstrate that ATDWA provides a stable and effective solution for addressing task imbalance in medical multi-task learning.

III. Methodology

A. Overall Architecture of Adaptive Weighting Framework

The model is developed within the SSC framework [7], which comprises a shared encoder and two task-specific heads for segmentation and classification. In this implementation, we replaces the original SSC encoder with a pretrained ResNet-34 backbone to enhance feature extraction while preserving the overall structure and training pipeline of the original framework.

ResNet-34 is selected because it has been extensively validated as an effective feature extractor in multi-task learning and medical image analysis [14, 15]. The shared backbone produces a unified feature representation for a given input image, which is simultaneously consumed by both prediction heads. During training, each task yields a task-specific loss that is passed to the Adaptive Temperature Dynamic Weight Averaging (ATDWA) module. ATDWA quantifies the temporal variability of each loss to derive dynamic task weights, which are then combined to form a single total loss. This total loss is used to update all model parameters via backpropagation. Integrating ATDWA into the SSC architecture preserves the benefits of shared representation learning while promoting more stable and adaptively balanced optimization across tasks. Fig. 1 illustrates the overall structure of the proposed multi-task framework.

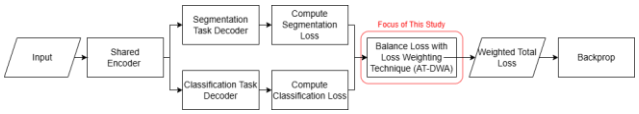

Figure 1. Overall Architecture

B. *Adaptive Temperature Dynamic Weight Average (Our Proposed Technique)*

The original DWA [4] employs a fixed temperature, which constrains its responsiveness to evolving task difficulty and inter-task divergence. In contrast, ATDWA augments DWA with a dynamic, uncertainty-aware temperature while retaining the same per-epoch weight update rule.

1) Loss Ratio

Let the set of tasks be denoted by $T = \{1, \dots, K\}$. For each task k, let $L_k(t)$ represent its loss at epoch t. The ratio loss $r_k(t)$ is computed as:

$$r_k(t) = \frac{L_k(t-1)}{L_k(t-2)+\epsilon} \tag{1}$$

with $\epsilon = 10^{-8}$ added for numerical stability.

ATDWA uses the same formulation as standard DWA [4] for computing the loss ratios $r_k(t)$. However, the subsequent formulas in (2) – (5) are proposed in this work.

2) Task-wise Temporal Uncertainty

Instead of computing uncertainty across tasks, ATDWA estimates temporal uncertainty for each task individually. For every task k, uncertainty is measured using the standard deviation of its loss values over the most recent window of three epochs (the window length can be set manually):

$$u_k(t) = StdDev(L_k(t), L_k(t-1), L_k(t-2)) \tag{2}$$

Higher values of $u_k(t)$ indicate that the task is fluctuating more strongly over time, signaling instability in its optimization trajectory.

To obtain a single uncertainty value that reflects the overall network stability, we compute the average across tasks:

$$\bar{u}(t) = \frac{1}{K}\sum_{k=1}^{K} u_k(t) \tag{3}$$

3) Adaptive Temperature Scaling

The temperature at epoch t is updated using the average temporal uncertainty:

$$T(t) = T_0(1 + \gamma\bar{u}(t)) \tag{4}$$

The base temperature T_0 is set to 2 by following the standard DWA formulation, where a temperature of 2 has been empirically shown to provide optimal and stable performance across a wide range of network architectures [6].

The sensitivity parameter γ is set to 0.8, determined through a grid search over the range of 0.1 to 0.9 with a step size of 0.1. The selected value yields the lowest standard deviation across all folds, indicating improved robustness to data variation.

When losses fluctuate strongly, $\bar{u}(t)$ increases, leading to a higher temperature and more stable weighting. When losses are relatively stable, the temperature remains close to its base value, allowing more decisive adjustments in task weights.

4) Adaptive Weight Update

Using the ratios $r_k(t)$ and the adaptive temperature $T(t)$, the weight for each task is computed as:

$$w_k(t) = \frac{\exp\left(\frac{r_k(t)}{T(t)}\right)}{\sum_{j=1}^{K}\exp\left(\frac{r_j(t)}{T(t)}\right)} \times \mathrm{K} \tag{5}$$

The multiplication by K ensures that the weights sum to K, matching the normalization form used in our implementation.

These weights are then applied to the corresponding task losses to compute the weighted total loss, which is subsequently used for backpropagation.

IV. Results and Discussion

A. *Experimental Settings*

The Breast Ultrasound Images (BUSI) dataset [23] used in this study contains 647 images, including 210 malignant cases. The dataset was split into 80% for training and 20% for testing. Figure 2 shows representative examples of the dataset. We adopted five-fold cross-validation on the training set, where each fold reserved 20% of the training data for validation. After applying data augmentation to the training split, the effective training set increased to approximately 2,900 images, while the validation set contained about 100 images per fold and the held-out test set remained at 129 images. This protocol helps mitigate overfitting and provides a more reliable estimate of model performance. For consistent input formatting, all images were resized to 256×256 pixels, and data augmentation was applied to increase input variability [19, 20]. Augmentations applied with a probability of 1.0 included grid distortion, random rotations of up to 45°, optical distortion, horizontal and vertical flips, and elastic transformations. These augmentations substantially

TABLE II. COMPARISON ON TESTING SET RESULTS BETWEEN ABLATION MODELS

Model	Segmentation						Classification			
	Acc	*Dice*	*IoU*	*AUC*	*Recall*	*Precision*	*Acc*	*F1-score*	*Recall*	*Precision*
SSC-ResNet	0.9576	0.7414	0.6502	**0.8887**	**0.7970**	0.7538	0.8961	0.8960	0.8961	0.8981
SSC-ResNet-GradNorm [6]	0.9570	0.7144	0.6291	0.8677	0.7498	0.7508	**0.9271**	**0.9271**	**0.9271**	**0.9284**
SSC-ResNet-DWA [4]	**0.9598**	0.7420	0.6565	0.8815	0.7789	0.7659	0.9039	0.9033	0.9039	0.9080
SSC-ResNet-ATDWA (Ours)	0.9589	**0.7561**	**0.6711**	0.8843	0.7837	**0.7816**	0.9178	0.9178	0.9178	0.9207

*Numbers in bold indicate the highest value

expanded diversity in the input space [19, 20], enabling the model to learn more robust and consistent representations.

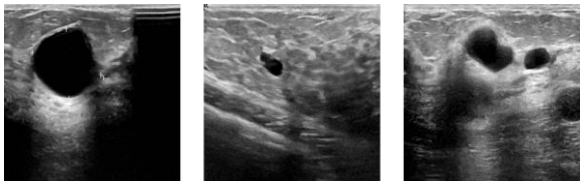

Figure 2. Samples of BUSI Dataset [23]

Model training was conducted using the Adam optimizer with a learning rate of 1×10^{-4}. A dropout rate of 0.3 was applied throughout the network to improve generalization. To mitigate training stagnation, the learning rate was reduced when the validation loss plateaued. Early stopping with a patience of 15 epochs was applied and monitored using the mean validation loss across the segmentation and classification tasks to maintain balanced optimization. A batch size of 16 was used for all methods, except for GradNorm [6], which required a reduced batch size of 8 due to its increased memory consumption for gradient-based weighting. Standard DWA [4] was implemented with a fixed temperature of T = 2, consistent with its original formulation.

We conduct a comparative study against prior studies and perform an ablation analysis involving multiple architectural and weighting configurations: SSC-ResNet, SSC-ResNet with GradNorm (SSC-ResNet-GradNorm), SSC-ResNet with DWA (SSC-ResNet-DWA), and the proposed SSC-ResNet with ATDWA (SSC-ResNet-ATDWA). To comprehensively assess performance, both segmentation and classification tasks are evaluated on the test set using widely adopted metrics in medical image analysis. Segmentation performance is evaluated using the Dice coefficient, IoU, segmentation accuracy (Acc), Recall, Precision, and AUC. Classification performance is evaluated using accuracy (Acc), F1-score, Recall, and Precision. All metrics adhere to their standard definitions, without task-specific modifications, ensuring a fair comparison across weighting strategies and consistency with prior literature. In addition to accuracy, we measure training memory usage and runtime of SSC-ResNet with and without adaptive weighting methods to evaluate their computational efficiency. Memory consumption is measured using PyTorch CUDA memory allocation statistics, and runtime is measured using PyTorch timing utilities [24].

B. Comparison with Previous Studies

Table I shows that the proposed SSC-ResNet-ATDWA achieves better overall performance than prior SSC-based work, including the baseline SSC [7]. Compared with SSC [7], our method yields substantial gains in segmentation metrics: the Dice and IoU scores increase by 5.4986% and 1.4863%, respectively, while the segmentation accuracy (Acc) decreases slightly by 0.5138%. Simultaneously, classification accuracy (Acc) and Recall improve by 5.8993%, accompanied by a 5.4230% increase in Precision. These results suggest that combining a pretrained ResNet-34 encoder with the ATDWA adaptive weighting mechanism improves both feature learning and task balancing, enabling our model to outperform prior methods while maintaining strong performance in segmentation and diagnosis.

TABLE I. COMPARATIVE PERFORMANCE WITH PREVIOUS STUDIES

Author	Segmentation			Classification		
	Acc	*Dice*	*IoU*	*Acc*	*Recall*	*Precision*
Byra et al. [22]	0.9270	0.6440	NA	0.8650	0.8200	NA
SSC [7]	**0.9639**	0.7167	0.6613	0.8667	0.8587	0.8733
SSC-ResNet-ATDWA (Ours)	0.9589	**0.7561**	**0.6711**	**0.9178**	**0.9178**	**0.9178**

*Numbers in bold indicate the highest value

C. Comparison on Ablation Models

This section presents the results of the ablation study on the test set. All metrics are averaged over five cross-validation folds to provide a robust estimate of generalization performance. Table II summarizes segmentation and classification outcomes across architectural and task-weighting variants. Additionally, qualitative comparisons are provided to assess lesion boundary consistency and illustrate the effects of different task-weighting strategies.

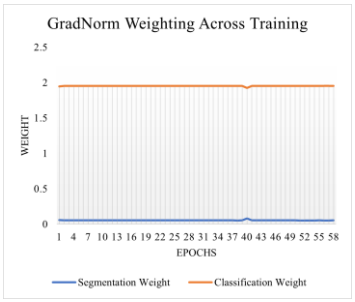

Figure 3. GradNorm loss weights across training epochs

Table II shows that GradNorm achieves the strongest classification performance, improving accuracy (Acc), F1-score, Recall, and Precision by 3.4602%, 3.4729%, 3.4602%, and 3.3713%, respectively, compared with SSC-ResNet. However, GradNorm substantially degrades segmentation performance, with decreases of 0.0616% in segmentation accuracy (Acc), 3.6447% in Dice, 3.2378% in IoU, 2.3642% in

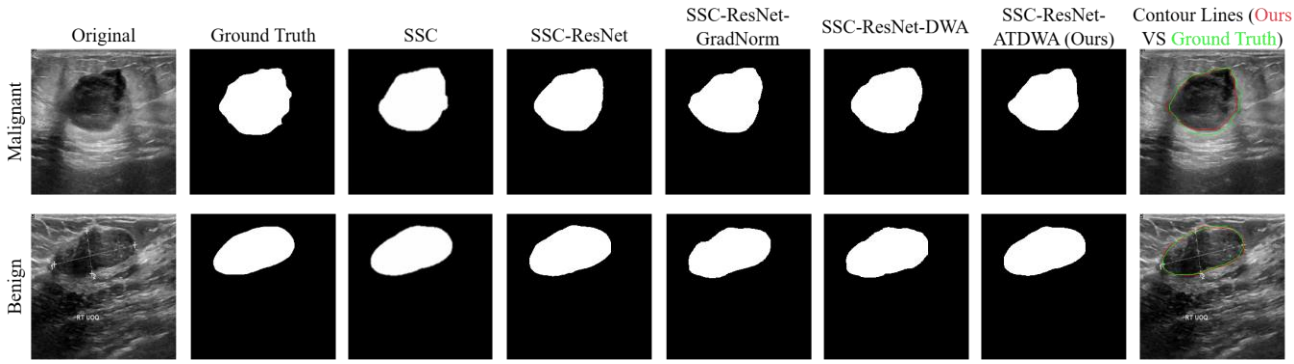

Figure 4. Comparison of Segmentation Results

AUC, 5.9178% in Recall, and 0.3950% in Precision. This inconsistency in GradNorm is driven by a rapidly emerging task imbalance during training, where GradNorm reallocates weights to prioritize the classification loss, causing the segmentation weight to approach zero, as shown in Fig. 3.

DWA exhibits a more balanced, though still imperfect, weighting behavior across tasks. Compared with SSC-ResNet, DWA improves segmentation accuracy (Acc), Dice, IoU, and Precision by 0.2269%, 0.0808%, 0.9678%, and 1.5999%, respectively. However, the segmentation AUC and Recall decreased by 0.8169% and 2.2640%. For classification, DWA yields modest gains, with accuracy (Acc), F1-score, Recall, and Precision increasing by 0.8651%, 0.8172%, 0.8651%, and 1.1054%, respectively

The proposed ATDWA achieves a more favorable balance between tasks by delivering consistent improvements in segmentation while remaining highly competitive in classification. Compared with SSC-ResNet, segmentation accuracy (Acc) increases by 0.1376%, which is only 0.0890% lower than DWA. Moreover, Dice, IoU, and Precision improve by 1.9866%, 3.2226%, and 3.6793%, respectively. The decreases in pixel-wise AUC and Recall are also notably smaller than those observed with GradNorm and DWA. For classification, ATDWA improves accuracy (Acc), F1-score, Recall, and Precision by 2.4221%, 2.4336%, 2.4221%, and 2.5111%, respectively, relative to SSC-ResNet. Compared with GradNorm, ATDWA consistently ranks as the second-best performing technique in classification, with reductions across all metrics remaining within approximately 1%. Unlike GradNorm, which substantially compromises segmentation, ATDWA preserves effective segmentation learning while achieving near-maximum classification performance. These results highlight the benefit of adaptive temperature adjustment in improving the stability and reliability of multi-task optimization.

Fig. 4 shows qualitative segmentation results for selected malignant and benign tumors. Across methods, the predicted masks generally capture the overall tumor morphology. However, none of the approaches delineates lesion boundaries, particularly for malignant cases where contours are often irregular and visually ambiguous. We visualize predictions using contour overlays to better assess the alignment of boundaries. While the proposed method reliably localizes lesion regions, boundary ambiguity remains a key challenge. On the right side of the malignant lesion, the subtle intensity difference between tumor tissue and background causes the predicted contour to contract inward, producing a narrower segmentation than the ground truth. At the left boundary, where background contrast is comparable to the lesion, the model shows a mild tendency to over-segment. Along the inferior boundary, the predicted contour follows a darker intensity transition, whereas the ground truth extends into brighter tissue regions consistent with tumor presence, suggesting that annotation ambiguity may also contribute to the discrepancy. Overall, these observations indicate that segmentation errors are largely driven by low-contrast and visually ambiguous regions, in addition to model limitations.

TABLE III. COMPUTATIONAL EFFICIENCY COMPARISON BETWEEN ABLATION MODELS

Model	Max Memory Usage (GB)	Longest Per Step Runtime (ms)	Longest Per Epoch Runtime (s)
SSC-ResNet	0.8046	179.5562	26.5903
SSC-ResNet-DWA [4]	0.8049	185.5686	26.0866
SSC-ResNet-GradNorm [6]	6.8029	1286.3955	258.8503
SSC-ResNet-ATDWA (Ours)	0.8049	187.4173	26.6796

Table III shows that the proposed ATDWA maintains a lightweight computational profile comparable to DWA and the baseline SSC-ResNet. ATDWA introduces only a slight increase in peak memory usage and training time relative to the unweighted model, with both step-level and epoch-level runtimes remaining close to those of DWA. In contrast, SSC-ResNet-GradNorm incurs substantial overhead, increasing peak memory usage by more than eight times and extending training time by approximately six to eight times compared with the baseline, DWA, and ATDWA. Overall, these results confirm that ATDWA provides adaptive task weighting while preserving training efficiency.

V. CONCLUSION

This study presents the SSC-ResNet-ATDWA, which adaptively modifies task weights according to temporal loss variability, showing enhanced performance in segmentation and classification tasks relative to previous SSC-based

approaches. We conducted a comprehensive evaluation against prior works and multiple ablation variants (SSC-ResNet, SSC-ResNet-GradNorm, and SSC-ResNet-DWA) using a five-fold cross-validation approach.

Compared to earlier SSC-based work, SSC-ResNet-ATDWA demonstrates significant improvements in key metrics, achieving increases of 5.4986% and 1.4863% in Dice and IoU scores, respectively, while accuracy decreased slightly by 0.5138%. For classification, accuracy (Acc) and Recall each increased by 5.8933%, accompanied by a 5.4230% increase in Precision. In comparison to the ablation models, ATDWA also achieved the strongest overall segmentation performance, improving Dice, IoU, and Precision by 1.9866%, 3.2226%, and 3.6793%, respectively, over SSC-ResNet, while maintaining segmentation accuracy (Acc), AUC, and Recall at levels comparable to those of DWA and SSC-ResNet. In classification tasks, ATDWA ranks just below GradNorm, with all metrics differing by about 1% from the top-performing model. However, ATDWA avoids the severe segmentation degradation caused by GradNorm's imbalance-driven gradient scaling, resulting in a more stable and reliable weighting strategy.

Future research may investigate the integration of validation-aware weighting decisions or the incorporation of gradient interaction modeling to improve optimization stability. Furthermore, assessing the proposed method on a broader range of multi-task learning benchmarks, beyond medical imaging and across diverse backbone architectures, would enhance the understanding of its general-purpose applicability.

ACKNOWLEDGEMENT

The authors appreciate financial support from the Institution of Research and Community Services, Universitas Atma Jaya Yogyakarta.

REFERENCES

[1] S. Chen, Y. Zhang, and Q. Yang, "Multi-Task Learning in Natural Language Processing: An Overview," ACM Comput. Surv., vol. 56, no. 12, pp. 1–32, July 2024, doi: 10.1145/3663363.

[2] Y. Zhang and Q. Yang, "A Survey on Multi-Task Learning," IEEE Trans. Knowl. Data Eng., vol. 34, no. 12, pp. 5586–5609, Dec. 2022, doi: 10.1109/tkde.2021.3070203.I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[3] S. Vandenhende, S. Georgoulis, L. Van Gool, D. Dai, M. Proesmans, and W. Van Gansbeke, "Multi-Task Learning for Dense Prediction Tasks: A Survey.," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 7, p. 1, Jan. 2021, doi: 10.1109/tpami.2021.3054719.

[4] S. Liu, E. Johns, and A. Davison, "End-to-End Multi-Task Learning with Attention," Mar. 28, 2018, Cornell University. doi: 10.48550/arxiv.1803.10704.

[5] A. Kendall, Y. Gal, and R. Cipolla, "Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics," May 19, 2017, Cornell University. doi: 10.48550/arxiv.1705.07115.

[6] Z. Chen, V. Badrinarayanan, C.-Y. Lee, and A. Rabinovich, "GradNorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multitask Networks," Nov. 07, 2017. doi: 10.48550/arxiv.1711.02257.

[7] C.-L. Yang, Y.-Y. Chen, and Y. Harjoseputro, "A hybrid approach of simultaneous segmentation and classification for medical image analysis," Multimed Tools Appl, May 2024, doi: 10.1007/s11042-024-19310-9.

[8] Y. Miao, Y. Tang, H. Ren, and J. Li, "The Impact of Integrating Shallow and Deep Information on Knowledge Distillation," *IEEE Access*, vol. 13, pp. 89831–89843, 2025, doi: https://doi.org/10.1109/access.2025.3571732.

[9] W. Wu, L. Huo, G. Yang, X. Liu, and H. Li, "Research into the Application of ResNet in Soil: A Review," *Agriculture*, vol. 15, no. 6, p. 661, Mar. 2025, doi: https://doi.org/10.3390/agriculture15060661.

[10] T. Nigussie, Y. Belayneh, B. Simegnew, Y. Y. Munaye, A. Molla, and W. Admass, "Brain Tumor Segmentation Using Multi-Task Learning," *2024 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*, pp. 31–36, Nov. 2024, doi: https://doi.org/10.1109/ict4da62874.2024.10777135.

[11] C. N, M. B, and S. B. R V, "Enhanced Image Classification Using Transfer Learning with ResNet50-V2: A Case Study on Wildlife Recognition," *2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*, pp. 1565–1570, Jan. 2025, doi: https://doi.org/10.1109/icmcsi64620.2025.10883359.

[12] H. Sun, I. Guyon, F. Mohr, and Hedi Tabia, "RRR-Net: Reusing, Reducing, and Recycling a Deep Backbone Network," *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–9, Jun. 2023, doi: https://doi.org/10.1109/ijcnn54540.2023.10191770.

[13] S. Divya, L. Padma Suresh, and A. John, "A Deep Transfer Learning framework for Multi Class Brain Tumor Classification using MRI," *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Dec. 2020, doi: https://doi.org/10.1109/icacccn51052.2020.9362908.

[14] P. Tschandl, C. Sinz, and H. Kittler, "Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation," *Computers in Biology and Medicine*, vol. 104, pp. 111–116, Jan. 2019, doi: https://doi.org/10.1016/j.compbiomed.2018.11.010.

[15] Lakshmi S, M. T. Somashekara, and Mohana Kumar S, "A Comparative Analysis of Texture and Deep Feature Extraction Techniques for Breast Cancer Diagnosis in Ultrasound Imaging," pp. 1–8, Aug. 2025, doi: https://doi.org/10.1109/nmitcon65824.2025.11187447.

[16] W. Quan and S. Liu, "Research on CAD Image Classification Method Based on Deep Neural Networks :Transfer Learning with ResNet Model," *2024 7th International Conference on Information Communication and Signal Processing (ICICSP)*, pp. 171–175, Sep. 2024, doi: https://doi.org/10.1109/icicsp62589.2024.10809285.

[17] S. Kim, T. Purdie, and C. Mcintosh, "Cross-Task Attention Network: Improving Multi-Task Learning for Medical Imaging Applications," Sept. 07, 2023. doi: 10.48550/arxiv.2309.03837.

[18] G. Dai et al., "Multi-Task Learning Network for Medical Image Analysis Guided by Lesion Regions and Spatial Relationships of Tissues," IEEE Trans. Circuits Syst. Video Technol., p. 1, Jan. 2025, doi: 10.1109/tcsvt.2025.3596803.

[19] Yang, C. L., Harjoseputro, Y., Chien, C. H., & Chen, Y. Y. (2025). Deep integration of conditional gan, attention mechanism, and image clustering for automated color separation and correction in textile screen printing. *Applied Intelligence*, *55*(11), 772.

[20] Yang, C. L., Cheng, M. C., Harjoseputro, Y., Chien, C. H., & Chen, Y. Y. (2024, June). Integrating generative adversarial networks and clustering techniques for automated screen printing. In *2024 6th International Conference on Computer Communication and the Internet (ICCCI)* (pp. 200-205). IEEE.

[21] Yang, C. L., Harjoseputro, Y., Hu, Y. C., & Chen, Y. Y. (2022). An Improved Transfer-Learning for Image-Based Species Classification of Protected Indonesians Birds. *Computers, Materials & Continua*, *73*(3).

[22] M. Byra *et al.*, "Joint segmentation and classification of breast masses based on ultrasound radio-frequency data and convolutional neural networks," *Ultrasonics*, vol. 121, pp. 106682–106682, Apr. 2022, doi: https://doi.org/10.1016/j.ultras.2021.106682.

[23] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in Brief*, vol. 28, p. 104863, Feb. 2020, doi: https://doi.org/10.1016/j.dib.2019.104863.

[24] Mateus Yonathan, "ORANGUTAN (Optimal Resource Allocation via Navigational Grouping, Utilization, Territorial Adaptation, and Negotiation): A Simulation Framework for Agent-Based GPU Resource Negotiation," *HAL (Le Centre pour la Communication Scientifique Directe)*, Aug. 2025, doi: https://doi.org/10.5281/zenodo.16989411.