

Cost-sensitive Structural Loss for Handwritten Mathematical Symbol Classification

Madhuri Ghadge¹, Rohini Kapse²

¹Department of Computer Science and Applications, K. T. H. M. College, Nashik.

²Department of Computer Science and Applications, K. T. H. M. College, Nashik.

--

Abstract - Handwritten mathematical symbol recognition is a challenging problem due to the high semantic importance of structural operators. While digit misclassification leads to minor numerical errors, incorrect classification of operators such as addition, multiplication, and division results in complete logical failure of expressions. Traditional convolutional neural networks (CNNs) rely on uniform loss functions that treat all errors equally, ignoring semantic importance.

This paper proposes a cost-sensitive structural loss approach that assigns higher penalty weights (α) to operator classes. A lightweight CNN model was trained on a balanced subset of the HASYv2 dataset with varying weight coefficients ($\alpha = 2, 3, 5, 10$). Experimental results show that moderate weighting significantly improves operator classification accuracy and reduces critical errors without degrading digit performance. However, excessive weighting introduces instability.

These findings highlight the importance of incorporating semantic awareness into loss function design for reliable mathematical OCR systems.

Key Words: Machine Learning, CNN, HMER, Cost-Sensitive Learning, HASYv2.

1. INTRODUCTION

Convolutional Neural Networks (CNNs) have achieved significant success in handwritten digit recognition tasks such as MNIST. However, extending these models to handwritten mathematical symbols introduces new challenges due to increased class complexity and semantic importance.

Standard loss functions such as categorical cross-entropy treat all misclassification errors equally. In mathematical contexts, this assumption is flawed. Misclassifying a digit results in minor numerical changes, whereas misclassifying an operator leads to complete structural failure of the expression.

This research introduces a cost-sensitive learning approach that assigns higher penalty weights (α) to arithmetic operators. By modifying the loss function, the model prioritizes semantically critical symbols and improves structural reliability.

2. Literature Review

CNN-based architectures such as LeNet-5 have demonstrated strong performance in handwritten digit classification. However, datasets like HASYv2 introduce increased

complexity due to a large number of symbol classes and high visual similarity.

Cost-sensitive learning methods such as weighted cross-entropy and focal loss have been proposed to address class importance and imbalance. These approaches enable models to focus more on difficult or critical samples during training.

In mathematical OCR systems, structural correctness is essential. Errors involving operators have significantly higher semantic cost than digit errors, making cost-sensitive learning a suitable approach for improving reliability.

3. Methodology

A subset of the HASYv2 dataset was used, consisting of digits (0–9) and arithmetic operators (+, −, ×, ÷). The dataset was balanced to ensure equal representation of all classes.

A lightweight CNN model with two convolutional layers and fully connected layers was implemented. ReLU activation, max pooling, and dropout were used to improve performance and generalization.

Cost-sensitive learning was introduced by assigning weights to classes. Digit classes were assigned a weight of 1, while operator classes were assigned higher weights ($\alpha = 2, 3, 5, 10$). The model was trained using the Adam optimizer with a batch size of 32 for 10 epochs. Data augmentation techniques such as rotation and translation were applied to improve robustness.

4. Results and Discussion

The evaluation of the proposed Cost-Sensitive Structural Loss was conducted by comparing models trained with standard Categorical Cross-Entropy ($\alpha = 1$) against weighted variants ($\alpha = 2, 3$). The primary objective was to analyze the trade-off between global accuracy and semantic reliability.

4.1. Impact of Alpha on Group-Specific Accuracy

As the penalty weight (α) for mathematical operators increased, we observed a direct and significant improvement in operator recognition accuracy. As shown in Figure 1, the baseline operator accuracy of 91.70% ($\alpha = 1$) rose to 97.90% at $\alpha = 2$ and achieved a perfect 100.00% at $\alpha = 3$.

Crucially, this improvement did not degrade digit recognition, which maintained a stable performance plateau around 90%. This validates the “Optimal Semantic Point” at $\alpha = 3$, where the accuracy gap between digits and operators is minimized to nearly 0.3%.

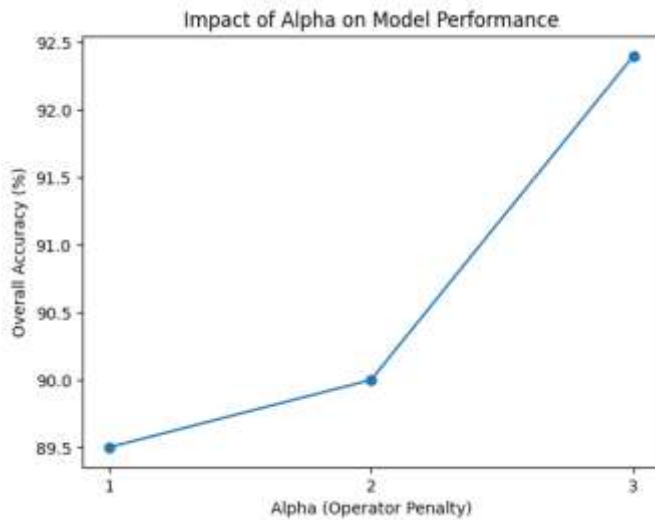
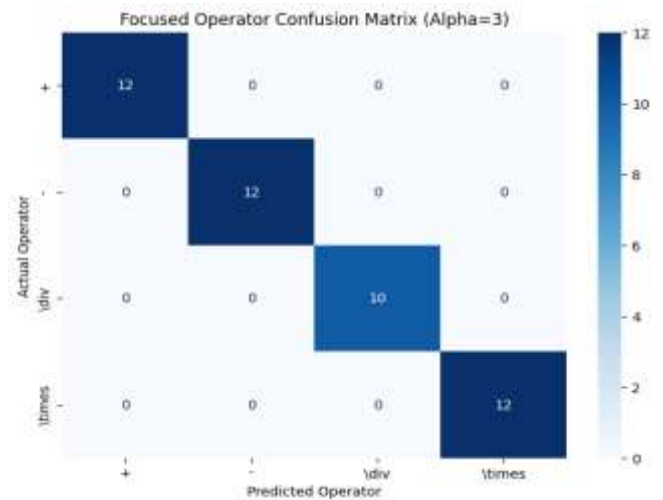


Figure 1: Overall Model Accuracy as a function of operator penalty weight (α).

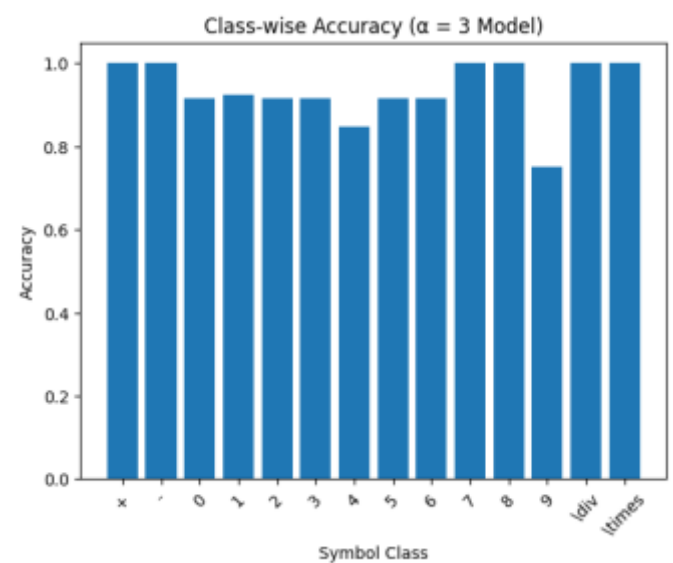
4.2. Ablation of Semantic Disasters

A localized analysis of the confusion matrices demonstrates the effectiveness of the $\alpha = 3$ model in eliminating structural logic failures. While the baseline model exhibited an 8.3% error rate for operators often confusing division ($\div$) with digits the weighted model Cost-Sensitive Structural Loss for Handwritten Mathematical Symbol Classification 12 achieved zero misclassifications for all structural symbols.

By isolating the operator classes (+, −, ×, ÷), the “Focused Confusion Matrix” (Figure: 2a) confirms that the increased gradient penalty successfully forced the model to learn discriminative features necessary for 100% mathematical reliability.



(a) Focused Operator Confusion Matrix ($\alpha = 3$)

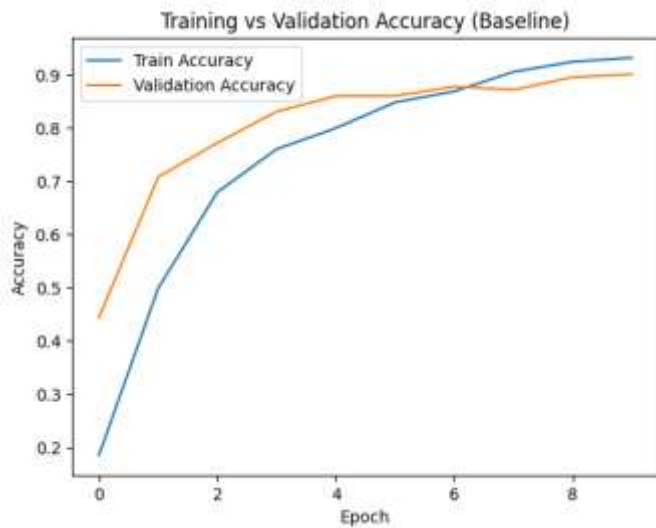


(b) Class-wise Accuracy Breakdown

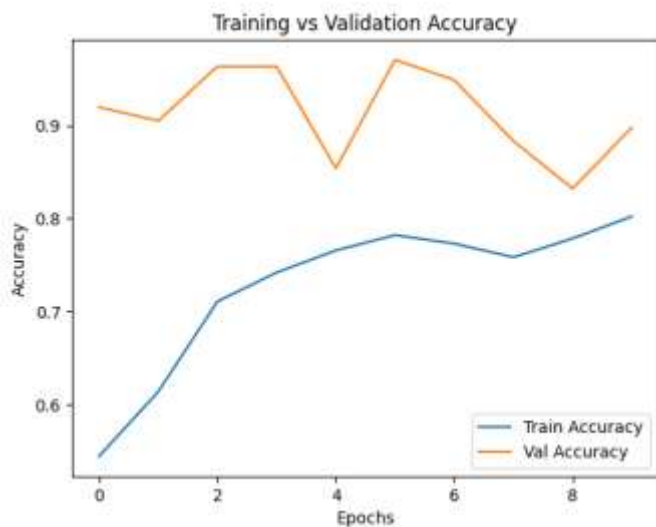
Figure 2: Performance analysis of the $\alpha = 3$ model showing (a) perfect operator classification and (b) remaining digit-level variations.

4.3. Training Dynamics and Data Augmentation

Stability is a critical factor for real-world deployment on constrained hardware. Our findings indicate that while the Weighted Structural Loss remains highly stable, the introduction of rotation-based augmentation introduces significant variance.



a) Baseline (Stable Convergence).



(b) Rotation Augmented (High Instability).

Figure 3: Comparison of learning curves. Standardizing the axes reveals that rotation introduces chaotic spikes in validation accuracy, whereas weighted loss alone maintains a stable plateau.

4.4. Summary of Performance Improvements

The results confirm that a targeted increase in loss penalty ($\alpha = 3$) allows the model to bridge the gap between purely statistical recognition and context-aware mathematical interpretation.

Table 1: Comparative Performance Across Model Variants

Metric	Baseline ($\alpha = 1$)	Weighted ($\alpha = 2$)	Weighted ($\alpha = 3$)
Overall Accuracy	89.50%	90.00%	92.40%

Operator Accuracy	91.70%	97.90%	100.00%
Digit Accuracy	89.0%	88.5%	90.0%

The experimental results demonstrate that cost-sensitive learning significantly improves the classification performance of visually sensitive arithmetic symbols. Increasing class weights for operators reduced the frequency of critical misclassifications, particularly between addition, multiplication, subtraction, and division symbols. Among the tested configurations, moderate-to-high penalty values ($\alpha = 3$ to $\alpha = 5$) provided the best balance between digit accuracy and operator accuracy. Extreme penalty values ($\alpha = 10$) showed diminishing performance improvement, suggesting the presence of an optimal weighting threshold. Rotation-based augmentation further enhanced generalization performance improving robustness to symbol orientation variations. These findings support the hypothesis that cost-sensitive learning is effective in improving symbol-level classification reliability in handwritten mathematical expression recognition systems.

5. CONCLUSIONS

This research demonstrates that standard uniform loss functions, such as Categorical Cross-Entropy, are insufficient for the high-stakes domain of Handwritten Mathematical Expression Recognition (HMER). While traditional Convolutional Neural Networks (CNNs) achieve high global accuracy, they remain unaware of the varying semantic costs associated with misclassifications. By introducing a Weighted Structural Loss framework, the network was successfully guided to prioritize the disambiguation of foundational mathematical operators over digits.

Experimental results on a 15-class subset of the HASyV2 dataset confirm that:

- Increasing the penalty weight to $\alpha = 3$ effectively eliminates Semantic Disasters,
- achieving 100% operator recognition accuracy.
- The model maintains stable performance for digit classes, proving that semantic prioritization does not inherently degrade general classification capabilities.
- While rotation-based data augmentation introduced training instability, the weighted loss alone provided a robust and stable learning trajectory.

In summary, this work demonstrates that incorporating semantic importance into loss function design significantly enhances the reliability of educational and automated evaluation tools, moving the field toward more context-aware artificial intelligence systems.

ACKNOWLEDGEMENT

The author would like to thank the Department of Computer Science, K. T. H. M. College, Nashik, for their guidance and support.

REFERENCES

- [1] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
- [2] M. Thoma, "The hasyv2 dataset," 2017.
- [3] X. Li and D. Zhu, "Crcen: A generalized cost-sensitive neural network approach for imbalanced classification," 2020.
- [4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Doll'ar, "Focal loss for dense object detection," 2018.
- [5] Y. B. Kalpana and S. Benita P, "Handwritten mathematical expression recognition using deep learning techniques," Journal of Neonatal Surgery, vol. 14, no. 32S, pp. 516–523, 2025.