A Comparative Study of Spatio-Temporal Segmentation Performance: AWS g4dn.xlarge vs. Google Colab T4 GPU

Pallab Mandal[1,2], Swaroop Srisailam[2], Praveen Pankajakshan[2], and Kumar Rajamani[2]
[1]Department of Mathematics and Scientific Computing, Indian Institute of Technology Kanpur, India
[2]Cropin AI Lab, Bangalore, India
Emails: pallabm22@iitk.ac.in, pallab.mandal@cropin.com, swaroop.srisailam@cropin.com,
praveenpankaj@ieee.org, kumar.rajamani@cropin.com

***Abstract*—In this study, we investigated the effect of training image Gaussian noise augmentation on the performance of semantic segmentation task in satellite images and bench-marked the results on two cloud computing platforms: Google Colab (T4-GPU) and AWS (g4dn.xlarge). Using a dataset of 100 each of training, validation, and test patches, we trained models for 15 epochs with a batch size of 2 (with and without Gaussian noise). The results showed notable performance differences between platforms. On the T4 GPU, the model achieved 72.96% accuracy and 31.00% IoU (Intersection over Union) without noise, slightly improving to 74.54% accuracy and 31.56% IoU with noise. In contrast, AWS g4dn.xlarge had better results, with 77.80% accuracy and 38.02% IoU without noise, and 79.15% accuracy and 38.51% IoU with noise. The significant variation, especially in IoU, suggests that hardware differences affect model performance. These findings highlight the importance of choosing appropriate computational resources in deep learning experiments. Future work will explore the relationship between hardware and training efficiency, and the theoretical understanding of random masking on constructing better learners.**

***Index Terms*—segmentation, Gaussian noise, satellite image, cloud computing platforms, model performance**

A. Motivation

Semantic segmentation of satellite imagery is essential for applications in land cover classification, environmental monitoring, and agriculture. Despite advancements in deep learning models, their performance heavily depends on data quality, volume, and computational resources. Recent studies suggest that noise-based data augmentation can improve model generalization. However, the effects of hardware variations on model performance, especially with noise augmentation, remain under-explored. This article examines the impact of Gaussian noise on segmentation performance and investigates how different hardware platforms influence these results.

B. Limitations of the State-of-the-Art Approaches

Prior work in semantic segmentation [1] focus on improving model architectures and using augmentations [2] like rotation, flipping, and scaling to improve accuracies. Augmenting the training images by adding noise received attention mostly in image classification, but its impact on pixel-level tasks like semantic segmentation is less clear. In addition, the influence of hardware on the performance of these techniques has never been explores.

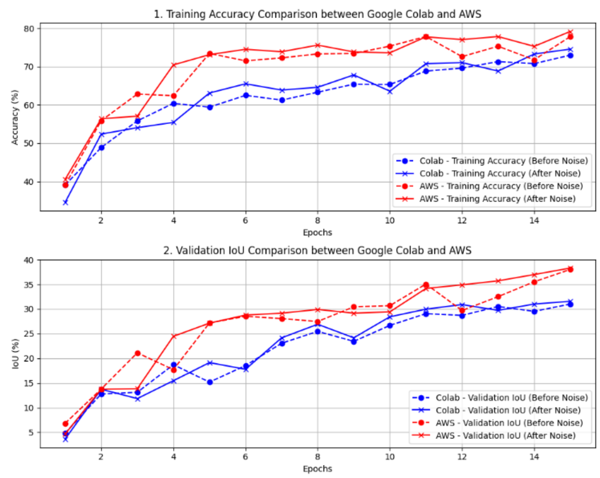

Fig. 1. Accuracy and IoU between Google Colab and the AWS instance

C. Performance Comparison between Google Colab and AWS

To evaluate the impact of hardware configurations on model performance, we conducted experiments on the Google Colab's T4 GPU and AWS's g4dn.xlarge instance with and without Gaussian Noise. Two key metrics are used: training accuracy and validation Intersection over Union (IoU). The results are given in Table I.

Metric	**T4 GPU**	**g4dn.xlarge**
Training Accuracy (Before Noise)	72.96%	77.80%
Training Accuracy (After Noise)	74.54%	79.15%
Validation IoU (Before Noise)	31.00%	38.02%
Validation IoU (After Noise)	31.56%	38.51%

TABLE I
COMPARISON OF TRAINING ACCURACY AND VALIDATION IOU BETWEEN GOOGLE COLAB'S T4 GPU AND AWS G4DN.XLARGE.

D. Related Work

A few recent studies demonstrate the effectiveness of machine learning in satellite image segmentation. For instance, [1] highlights how CNNs improve segmentation accuracy for agricultural monitoring, while [2] shows that Gaussian noise augments model generalization. Research on cloud platforms like Google Colab and AWS emphasizes how hardware differences impact training efficiency [3]–[6]. These findings highlight the need to optimize model architectures and training setups to enhance segmentation outcomes in remote sensing.

E. Key Insights and Contributions

This work shows that Gaussian noise augmentation, combined with different hardware setups, significantly affects semantic segmentation performance. Our experiments revealed that noise can improve accuracy and IoU, but these improvements tend to vary depending on the platform. Models trained on the g4dn.xlarge outperformed T4-GPU, and hence hardware must also be considered during model evaluation. Our key contribution is providing empirical evidence of the combined effects of noise augmentation and the choice of hardware.

F. Methodology

In our experiments, Gaussian noise from a standard normal distribution $\mathcal{N}(0,1)$ was added on 25% of the dataset during the pre-processing stage to assess model robustness. Noise was applied to each sample in the first dimension of the 4-D image data ($T \times C \times H \times W$), preserving spatial structure. The noisy images, denoted by $\mathbf{y}$, were generated as $\mathbf{y} \approx \mathbf{x} + \mathbf{n}$, where $\mathbf{n} \sim \mathcal{N}(0,1)$. One of the hypothesis is that the subsequent addition of noise, can lead to random masking (similar to masked autoencoders) of the image and by reconstructing the image, the models can become better learners.

G. PASTIS Dataset and the U-TAE Model Overview

The PASTIS dataset [7] comprises $2,433$ multispectral image sequences ($10 \times 128 \times 128$) from Sentinel-2, collected between September 2018 and November 2019. Each sequence contains 38 to 61 time observations, spanning over 4000 km^2 of diverse French regions with a $10-$meter spatial resolution. About 28% of the images have partial cloud cover. The dataset provides semantic and panoptic annotations to support land-use mapping and agricultural time series analysis. The details on the U-TAE model (U-Net with Temporal Attention Encoder) architecture used in processing these sequences are given in [7].

H. Reasons for Performance Discrepancy between AWS g4dn.xlarge and Google Colab T4 GPU

Several factors might have contributed to the observed differences in accuracy and Intersection over Union (IoU).

- **Hardware Specifications**: The AWS g4dn.xlarge instance features NVIDIA T4 GPUs with higher memory bandwidth and optimized Tensor Core performance, potentially allowing for faster training and inference.
- **Batch Size**: A smaller batch size of 2 used by both environments may affect convergence rates.
- **Thermal Throttling**: Google Colab may experience thermal throttling on shared resources, impacting performance during prolonged training sessions, whereas the AWS instance maintains consistent levels.
- **Data Loading and Preprocessing Speed**: Variations in data loading and preprocessing speeds could lead to differences in effective training time, with AWS potentially offering optimized data handling.

Properties	**g4dn.xlarge GPU**	**T4 GPU**
Performance		
Tensor Cores	320	320
CUDA Cores	$2,560$	$2,560$
Memory		
Memory Capacity	16 GB	16 GB GDDR6
General Specifications		
GPU Architecture	Tensor Core	Turing Architecture

TABLE II
COMPARISON OF AWS G4DN.XLARGE GPU AND NVIDIA T4 GPU SPECIFICATIONS

I. Conclusion

In this study, we analyzed the effect of Gaussian noise on segmentation models trained with the PASTIS dataset. Gaussian noise enhanced model robustness (similar effect as masking), yielding improved accuracy and IoU metrics on both Google Colab's T4 GPU and AWS g4dn.xlarge, with the latter consistently outperforming due to superior hardware and resource stability.

These findings underscore the importance of evaluating model performance across platforms, and random noise addition as masked learners. Future work should explore optimized architectures and additional augmentation techniques.

REFERENCES

[1] J. Choudhury and S. S. Gill, "Performance analysis of machine learning models for data visualisation in sme: Google cloud vs. aws cloud," in *Applications of AI for Interdisciplinary Research*. CRC Press, 2025, pp. 171–184.

[2] H. Kimm, I. Paik, and H. Kimm, "Performance comparision of tpu, gpu, cpu on google colaboratory over distributed deep learning," in *2021 IEEE 14th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC)*. IEEE, 2021, pp. 312–319.

[3] R. Narayanam, "Accelerating deep learning using gpus," *International Research Journal Of Engineering and Technology (IRJET)*, vol. 8, no. 3, pp. 1592–1600, 2021.

[4] V. Sharma, G. K. Gupta, and M. Gupta, "Performance benchmarking of gpu and tpu on google colaboratory for convolutional neural network," in *Applications of Artificial Intelligence in Engineering: Proceedings of First Global Conference on Artificial Intelligence and Applications (GCAIA 2020)*. Springer, 2021, pp. 639–646.

[5] Y. S. Voon, Y. Wu, X. Lin, and K. Siddique, "Performance analysis of cpu, gpu and tpu for deep learning applications," *Professor Ka Lok Man, Xi'an Jiaotong-Liverpool University, China Professor Young B. Park, Dankook University, Korea Chairs of CICET*, vol. 16, p. 12, 2021.

[6] X. Zhang, T. Zhang, J. Lu, X. Fu, and F. Reveriano, "The effect of high performance computer on deep neural network," *Engineered Science*, vol. 15, pp. 67–79, 2021.

[7] V. S. F. Garnot and L. Landrieu, "Panoptic segmentation of satellite image time series with convolutional temporal attention networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 4872–4881.