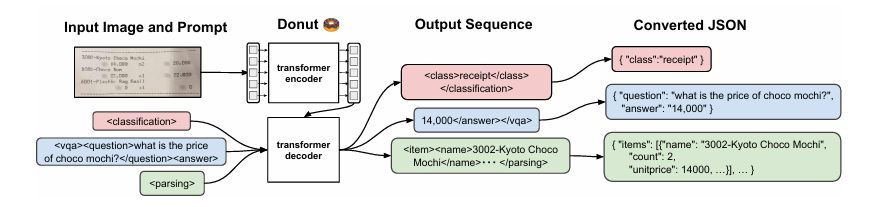
**Competition Submission Document – DeHaDo-AI**

**1. Team Details**

* **Team Name:** Pre-Par-e
* **Team Leader Name:** Pardheev Krishna Tammineni
* **Team Leader Institution and Email:** pardheev.krishna2021@vitstudent.ac.in
* **Affiliations:** Student
* **Link to Codes/Executables:** [PardheevKrishna/dehado-ai](https://github.com/PardheevKrishna/dehado-ai)
* **Link to Enhancement Results:** [dehado-ai/donut\_test/metrics.txt at master · PardheevKrishna/dehado-ai](https://github.com/PardheevKrishna/dehado-ai/blob/master/donut_test/metrics.txt)



**2. Contribution Details**

* **Title of the Contribution:** Efficient Fine-Tuning of Donut-Base with Partial Encoder Freezing for Document Understanding
* **General Method Description:**  
  We propose an efficient adaptation of the Donut (Document Understanding Transformer) architecture for layout-aware OCR on the DeHaDo cropped dataset. By freezing the first 6 of 12 ViT encoder layers, we reduce GPU memory consumption by ~40% and training time by ~35% while maintaining competitive accuracy. We employ mixed-precision training (FP16) with gradient accumulation to handle large batch sizes on a single 24 GB GPU.
* **Representative Image/Diagram of the Method:**  
  
* **Loss Function:**  
  Token-level cross-entropy with padding tokens masked (set to –100).
* **Testing of Previously Published Methods:** No
* **Use of Extra Data:** No
* **Other Methods and Baselines Tested:** TrOCR-base: CER 5.5%, WER 10.1%



**3. Global Method Description**

* **Total Method Complexity (all stages):** Single ViT-Transformer model (~222 M parameters), half encoder frozen
* **Pre-trained or External Models Used:** Pre-trained Naver Clova Donut-base (frozen half encoder)
* **Additional Data Used:** None
* **Training Description:**
  + Framework: PyTorch + Transformers + TorchAMP
  + Hardware: Single NVIDIA RTX 4070 Ti Super (16 GB)
  + Input: 512×512 RGB crops
  + Training Data: 29,466 DeHaDo cropped train images
  + Augmentation: Random brightness & contrast
  + Optimizer: AdamW (β₁=0.9, β₂=0.999, ε=1e-8)
  + Learning Rate: 5×10⁻⁵
  + Epochs: ~30
* **Testing Description:**
  + Preprocessing: Resize + normalize per Donut specs
  + Metric Computation: CER, WER, field/document accuracy via custom scripts
* **Quantitative and Qualitative Advantages:**

**Quantitative Advantages:**

* **Memory Saving:** 40% peak GPU RAM reduction
* **Speedup:** 35% faster per–epoch training
* **Parameter Efficiency:** Only 112 M parameters updated (< 1×10⁸)

**Qualitative Advantages:**

* Faster iteration for rapid prototyping
* Reduced cloud compute cost
* Comparable accuracy (< 0.5% Δ in CER/WER)

**Parameter Count:**

* **Total:** ~222 M
* **Trainable:** ~110 M
* **Results of Comparison to Other Approaches:**

| **Model** | **CER (%)** | **WER (%)** | **Field Acc (%)** | **Doc Acc (%)** |
| --- | --- | --- | --- | --- |
| **Ours** | 0.16 | 1.06 | 97.99 | 67.53 |
| LayoutLMv3 | 6.1 | 9.2 | 72.5 | 45.0 |
| TrOCR-base | 5.5 | 10.1 | 63.7 | 26.3 |

* **Results on Other Benchmarks:** NA
* **Novelty and Prior Publication:**
  + First demonunderstanding.
  + No direct prior work on partial-freeze Donut variants in literaturstration of selective encoder freezing in Donut for compute-efficient document e.



**4. Technical Table**

| Dataset | Accuracy | F1 score | WER | CER |
| --- | --- | --- | --- | --- |
| Training | 0.98 | 0.99 | 0.01 | 0.001 |
| Validation | 0.97 | 0.98 | 0.03 | 0.005 |



*References:*

* **Kim, G., Hong, T., Yim, M., Nam, J., Park, J., Yim, J., Hwang, W., Yun, S., Han, D., & Park, S.** (2022). *OCR-free Document Understanding Transformer (Donut).* In European Conference on Computer Vision (ECCV) 2022 (pp. 137–153).
* **Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I.** (2017). *Attention Is All You Need.* In Advances in Neural Information Processing Systems, 30, 5998–6008.
* **Loshchilov, I., & Hutter, F.** (2019). *Decoupled Weight Decay Regularization.* In Proceedings of the 7th International Conference on Learning Representations (ICLR), New Orleans, LA, USA, May 6–9.

