This report is a continuation of the work presented in Part-1. This also means that we have relied on our own previously developed solutions to facilitate the execution of Part-2

Training Description

A curriculum highlights the importance of using data in a developmentally structured order (youngest to eldest) to train deep network models effectively. As noted in the previous report, infant vision parameters mature progressively with age. Training the model with data in this natural sequence not only mimics this behaviour, but could also potentially outperform random data sequences[2]. Our project explores training of different networks under different transformations in a developmentally progressive manner for the Classification Task, as illustrated in ***Fig 1***. This work utilises EfficientNet-B2[1] as the training model due to its well-balanced tradeoff between size and accuracy. The classifier layer of the EfficientNet-B2 has been replaced by a custom fully connected layer and includes a dropout layer with rate 0.2 to prevent overfitting. Mixed precision training was used for faster training while reducing memory usage on GPUs. Additionally garbage collector and emptying of the GPU cache was done after every epoch to remove unnecessary data and prevent any memory leaks. A training checkpoint (.pth files) is saved after every last epoch in the sequence. These '.pth' files include epoch count, model state, optimiser state and the evaluation metrics. Timing was monitored using the *timeit* module to estimate efficiency and compare stages.

Model Overview	**Training Details**	**Compute Setup**
Model: EfficientNet-B2 ***Parameters:*** 9.1M	***Dataset:*** Subset of Tiny-Imagenet-200 (10 classes, 500 images per class) ***Loss Function:*** Cross-Entropy Loss ***Optimizer:*** Adam (learning rate 0.01) ***Batch Size:*** 64 (with shuffling)	***Hardware:*** Tesla T4 GPU ***Platform:*** Google Colab

Table1: Training description

Model	**Transform**	**Epochs**
1	No transforms (model 1)	65 epochs
2	Visual Acuity Curriculum (model 2)	10 epochs: Age 1 months 15 epochs: Age 5 months 40 epochs: Age 13 months
3	Contrast Sensitivity Curriculum (model 3)	10 epochs: Age 3 months 15 epochs: Age 7 months 40 epochs: Age 13 months
4	Contrast Sensitivity and Visual Acuity Shuffle Curriculum (model 4)	5 epochs: Age 3 months (Contrast Sensitivity) 10 epochs: Age 3 months (Visual Acuity) 20 epochs: Age 13 months (Contrast Sensitivity) 30 epochs: Age 13 months (Visual Acuity)

Table 2: Training conditions

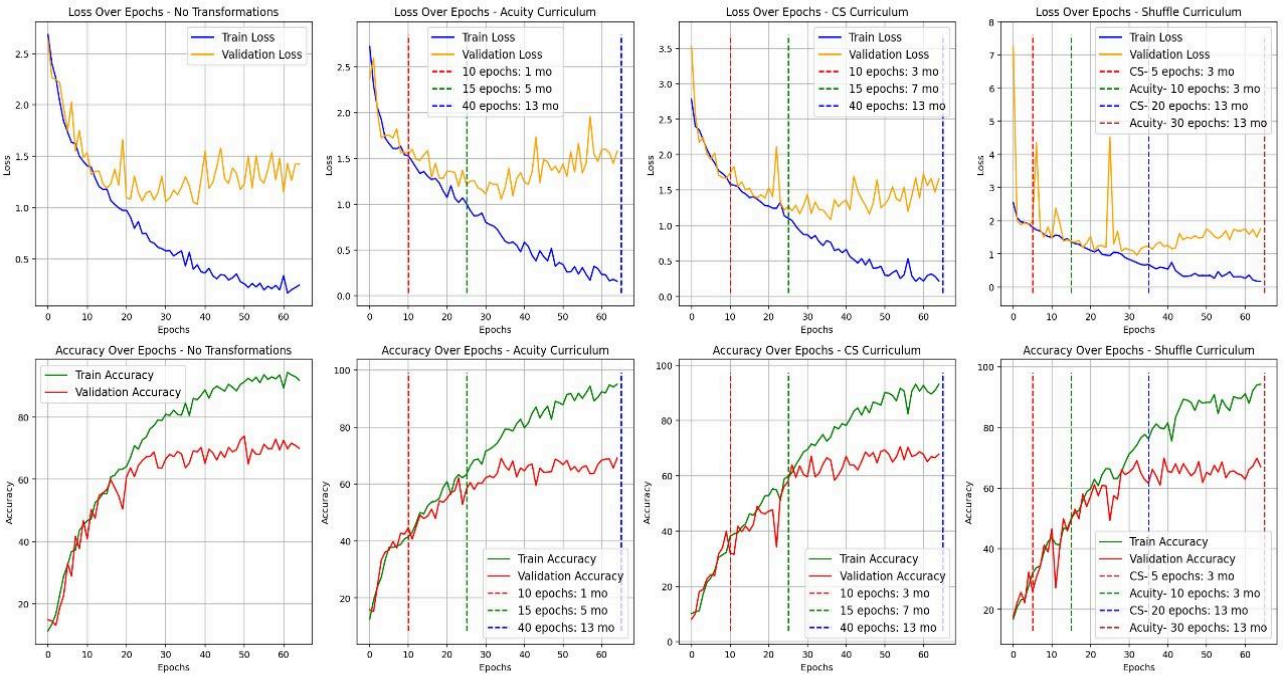

Figure 1: Variations of loss and accuracy over epochs for the following models - Model 1*(top left and bottom left)*; Model 2*(top middle-left and bottom middle-left)*; Model 3*(top middle-right and bottom middle-right)*; Model 4*(top right and bottom right)*.

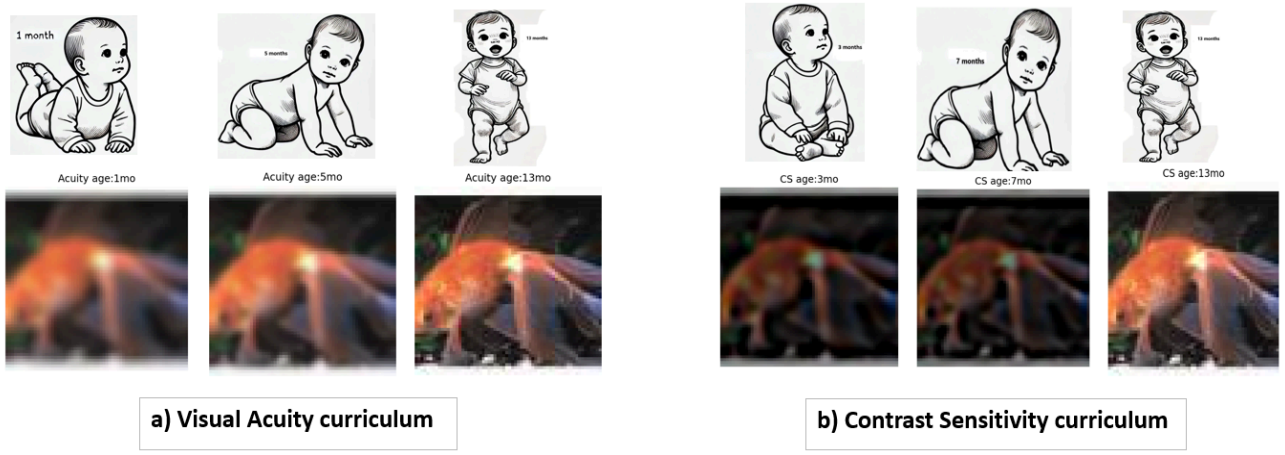

Figure 2: Visualizing curriculum (a) Visual Acuity: 1 month, 5 months, 13 months; (b) Contrast Sensitivity: 3 months, 7 months, 13 months.

---

[1] Tan, M., & Le, Q.V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *ArXiv, abs/1905.11946*, https://doi.org/10.48550/arXiv.1905.11946

[2] Saber Sheybani, Himanshu Hansaria, Justin N. Wood, Linda B. Smith, and Zoran Tiganj. (2024). Curriculum learning with infant egocentric videos