Project Report 2

Submitted by: Umang Tripathi, Raj Hemant Panchal, Sanal Varol

In this part of the project, we have applied image transformations, namely visual acuity and contrast sensitivity, to the TinyImageNet dataset and have trained them on a CustomResNet18 model using curriculum learning. The following sections discuss the experimental settings, the adopted curriculum, and the visualized results.

Selection of previous solution

Our solution in the previous task was primarily focused on grayscale images. When extended to RGB images, the contrast sensitivity transformation did not produce satisfactory results. Therefore, we have adopted the solution proposed by Do and Naik.

Experimental settings

Dataset - The TinyImageNet [1] dataset, consisting of 100k images across 200 classes (500 images per class) downsized to 64×64 RGB images is used for the multiclass classification task. Image transformations corresponding to four age groups (1, 6, 12, and 24 months) are applied on the complete dataset. Consequently, the total number of training and test images are 320k and 80k, respectively. A batch size of 64 is used.

Deep learning architecture used - We utilize a CustomResNet18 model for this task. The original ResNet18 [2] consists of 5 convolutional blocks. However, we use only the first 3 blocks, followed by a global average pooling layer, to avoid overfitting, as the image size (64×64) is much smaller than the standard 224×224. Additionally, we reduce the kernel size and padding in the first convolutional block to 5 and 1, respectively, for better feature extraction. The extracted feature map is flattened, and a fully connected (FC) layer with 128 filters maps it to the desired number of output classes.

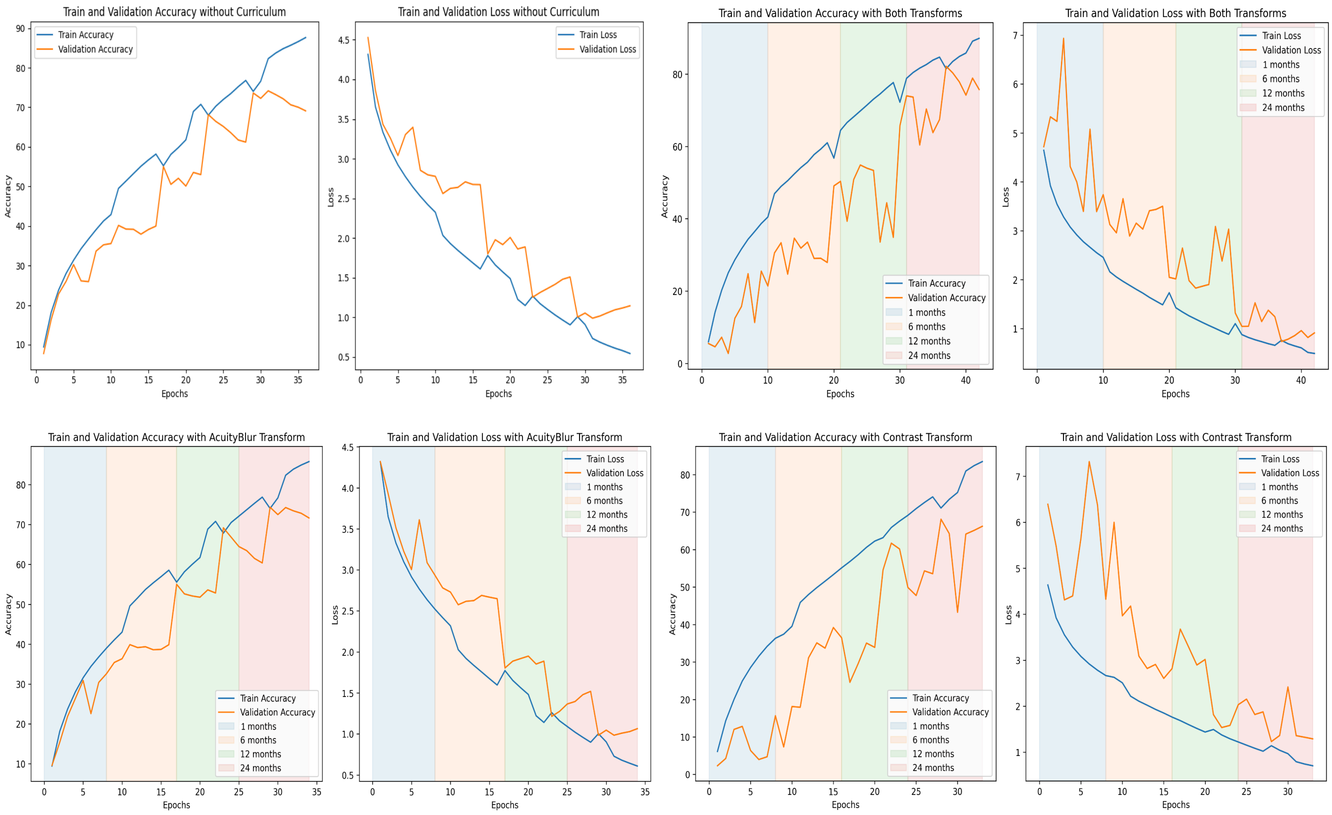
Hyperparameters - We train four different models (without curriculum, with visual acuity, with contrast sensitivity, and with both transformations) for 128 epochs each, divided into age-specific phases, with each age group trained for 32 epochs. We use the Adam optimizer with a learning rate of 0.001, beta values of (0.9, 0.999), and a weight decay of 1e-4 to prevent overfitting. A StepLR scheduler reduces the learning rate by a factor of 0.5 every 10 epochs to stabilize training in the later stages. We adopt the Cross-Entropy Loss function as the optimization criterion. The entire training process is carried out on NVIDIA TESLA P100 GPUs on Kaggle.

Curriculum description

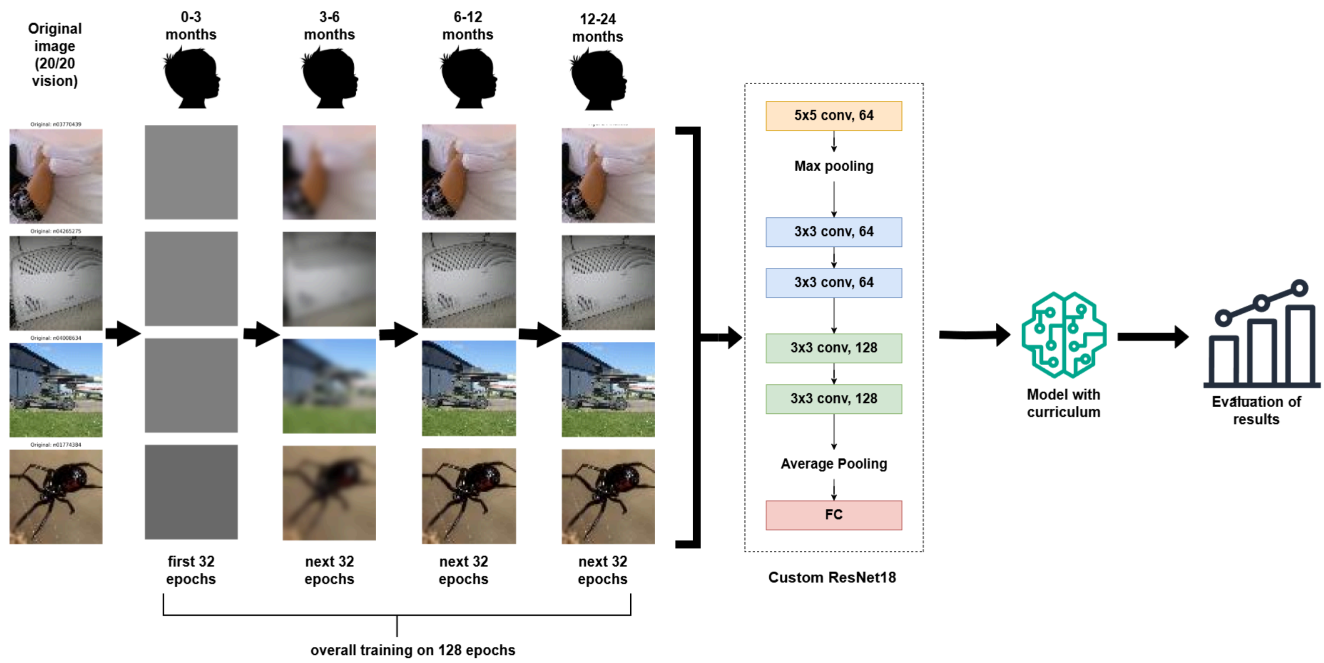
We employ a curriculum learning strategy, dividing the training process into sequential phases corresponding to different age groups. Curriculum learning [3], inspired by human education, gradually introduces tasks of increasing complexity or specificity, facilitating better learning progression. In this experiment, we expose the model to datasets representing different age groups, with each group trained for 32 epochs. At the start of each phase, we update the dataset and dataloaders to focus on the next age group.

Each age group's training includes an early stopping mechanism with a patience of 5 epochs to prevent overfitting. We save the model's parameters based on the best validation loss achieved during each phase. If no improvement is observed for 5 consecutive epochs, training for that age group terminates early. By gradually introducing age-specific data, the model systematically learns developmental visual patterns across age groups. At the end of the training process, we load the model with the weights corresponding to the lowest validation loss recorded during the entire process.

Learning curves



Curriculum visualization figure



References

[1] Le, Ya and Xuan S. Yang. "*Tiny ImageNet Visual Recognition Challenge.*" (2015).

[2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[3] Saber Sheybani, Himanshu Hansaria, Justin N. Wood, Linda B. Smith, and Zoran Tiganj. 2024. Curriculum learning with infant egocentric videos. *In Proceedings of the 37th International Conference on Neural Information Processing Systems (NIPS '23)*. Curran Associates Inc., Red Hook, NY, USA, Article 2360, 54199–54212.