Model Card: DenseNet121

1 Model Overview

Architecture: DenseNet121 builds dense connections between layers within each block, allowing each layer to receive input from all preceding layers. This design encourages feature reuse and improves parameter efficiency, which contributes to its competitive performance.

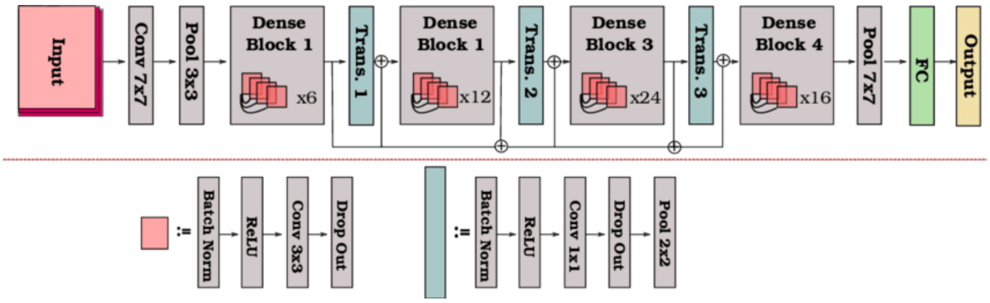

Figure 1: DenseNet121 Architecture [1]

Training Data: The model is trained on the ASL Alphabet dataset containing 87,000 images resized to 200x200 pixels, spanning 29 classes (26 letters plus 3 additional classes to aid live classification). 10 percent is held out as a test set, so 78,300 images are used in training

Table 1: Overview of Data Split and Image Specifications

Category	Number of Images	Percentage
Overall Dataset	87,000	100%
Training Set	78,300	90%
Training (90%)	62,640	72%
Validation (10%)	15,660	18%
Test Set	8,700	10%

Hyperparameters:

- Optimizer: Adam
- Loss Function: Categorical Cross-Entropy
- Learning Rate: 0.01
- Batch Size: 64

- Epochs: 20 (best validation accuracy is chosen)

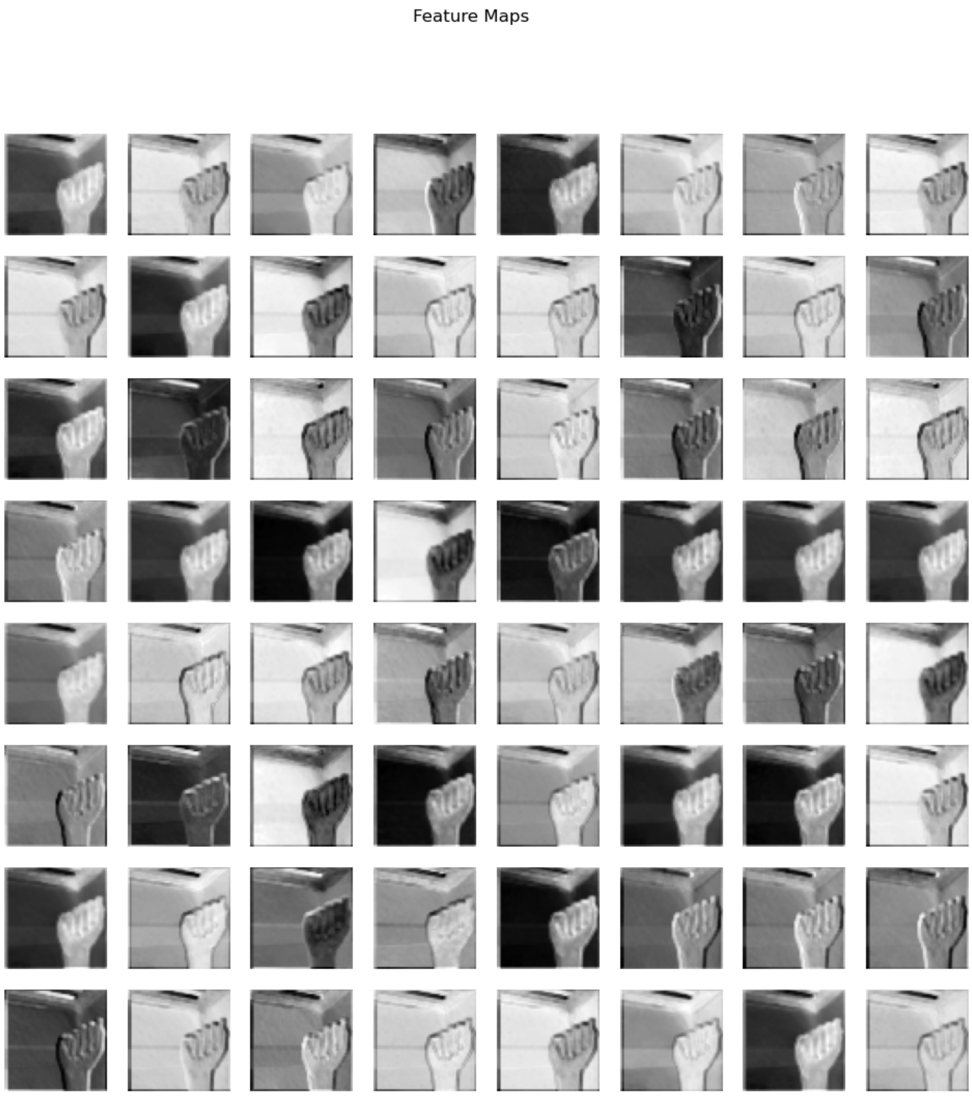

Figure 2: DenseNet121 Feature Maps

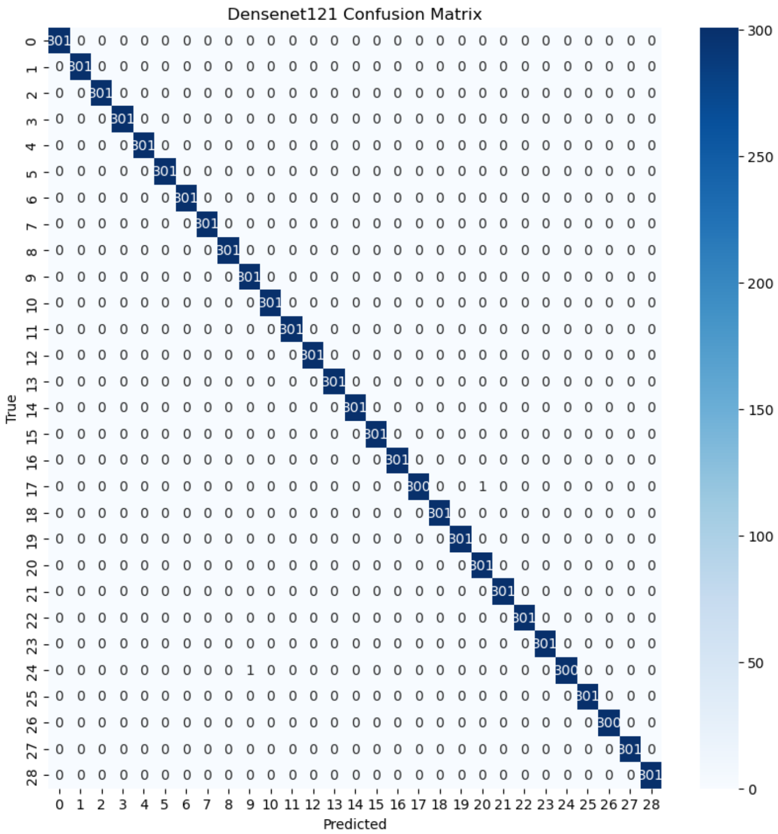

Figure 3: Confusion Matrix for DenseNet121

2 Intended Use

DenseNet121 is applied to image classification for American Sign Language (ASL) alphabet letters using a pre-trained model and fine-tuning through transfer learning. The model is intended for:

- Classifying ASL alphabet hand signs from static images
- Real-time sign language interpretation using live video inputs

3 Performance

On an independent test dataset, DenseNet121 achieved:

- **Test Accuracy:** 99.98%
- **Precision:** 1.0
- **Recall:** 1.0

DenseNet121's high performance underscores its capability for near-perfect classification, making it well-suited for both offline and live video recognition tasks.

4 Limitations

While achieving high accuracy, DenseNet121 has the following limitations:

- Sensitivity to variations in image quality and lighting conditions during live classification
- Dependence on precise hand positioning for accurate real-time classification
- Practical challenges when deploying on resource-constrained devices

5 Ethical Considerations

- Ensure the ASL dataset is diverse and representative to avoid biased outcomes
- Implement privacy protections when deploying real-time video classification systems

References

[1] "Leveraging sparse and dense features for reliable state estimation in urban environments," https://www.researchgate.net/publication/334170752_Leveraging_Sparse_and_Dense_Features_for_Reliable_State_Estimation_in_Urban_Environments, accessed: 2023-10-07.