Resume Classification using deep learning

Objective:

The project aims to classify images into "Resume" and "Non-Resume" categories using visual features with the help of deep learning techniques for image classification.

Dataset Details:

Data Sources: Utilized Roboflow as a primary data collection source. Collected approximately 150 images for both resume and non-resume classes. The distribution of classes are

- Resume Images: 150
- Non-Resume Images: 150

{Newspapers: 30, Bills and Personal Collection (Phone): 30, Open Source Research Papers: ~90}

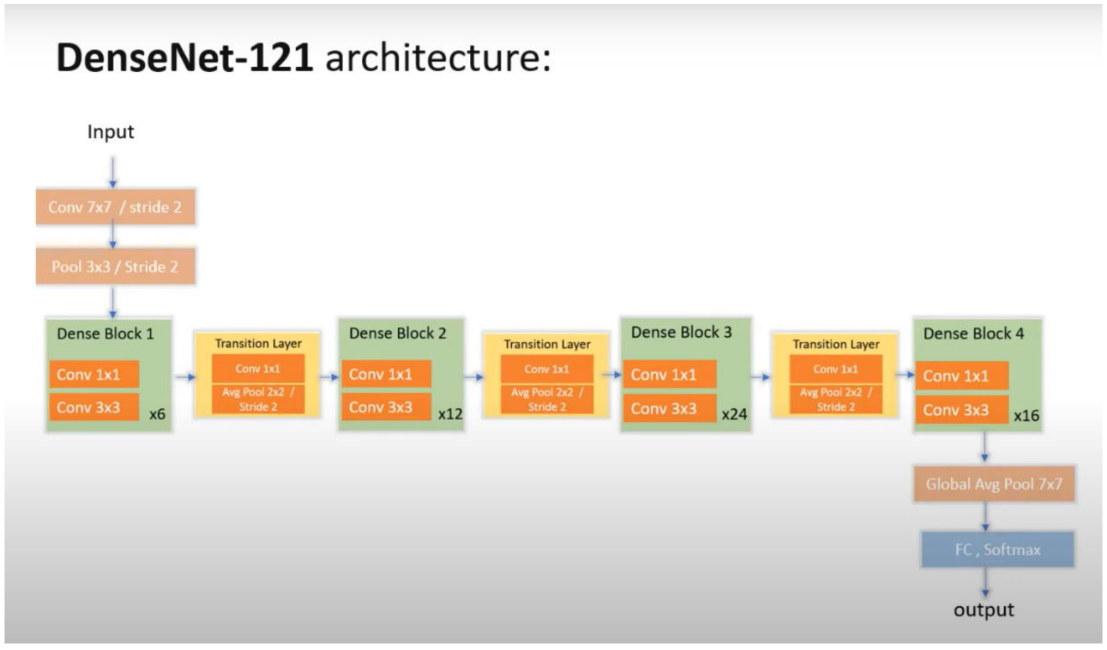
Data Augmentation:

Implemented data augmentation using the TensorFlow ImageDataGenerator with various transformations, such as rotation, width and height shifts, shear, zoom, brightness adjustments, and horizontal flipping.

Augmented each class to have a total of 300 images (original + augmented). The augmented dataset enhances the model's robustness and generalization to various visual characteristics.

Model Architecture:

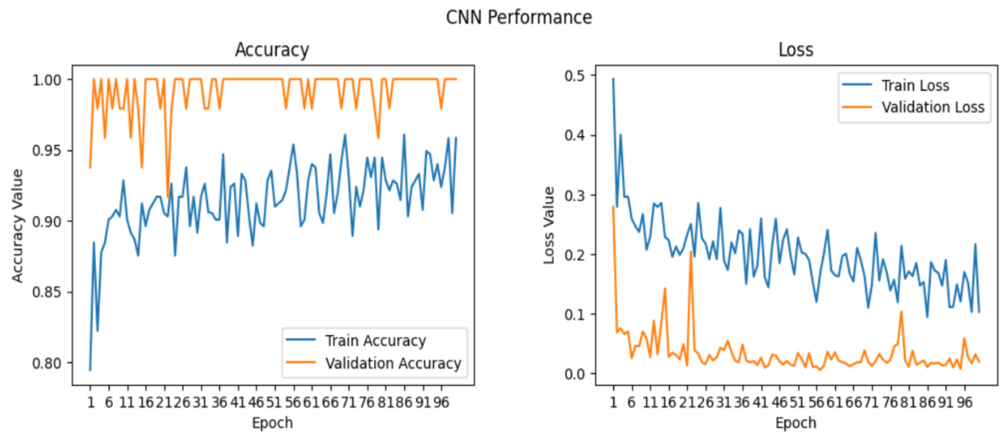
Initially, I implemented a basic CNN model, achieving a respectable accuracy of 87.6% on the test dataset. Subsequently, I adopted a more sophisticated approach by leveraging the DenseNet121 model with pre-trained weights from ImageNet, resulting in exceptional performance with 100% accuracy on the test data.

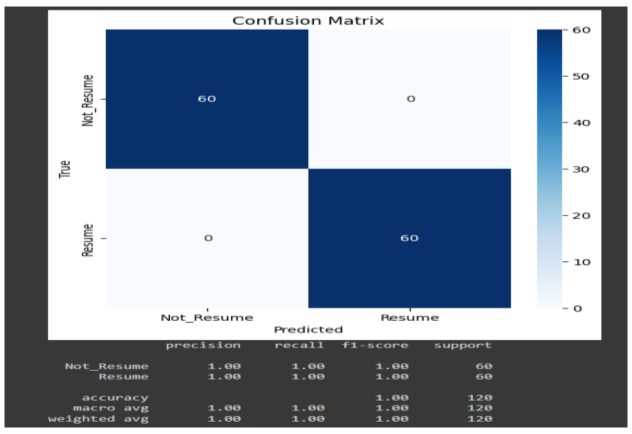


- DenseNet121's densely connected blocks enhance feature learning by promoting information flow across layers, crucial for capturing intricate visual patterns in resume images.
- Leveraging pre-training on ImageNet, the model initializes with diverse visual representations, aiding resume classification by recognizing relevant features even with a limited dataset.
- Global Average Pooling (GAP) in the architecture focuses on the global context of features, enhancing the model's ability to make precise decisions based on overall visual cues.
- In summary, DenseNet121's dense connectivity, transfer learning capabilities, and thoughtful architectural choices like GAP and sigmoid activation collectively make it well-suited for resume classification

Training Strategy:

After data augmentation, the dataset for each class expanded to 300 images. To evaluate the model, a test set of 60 images per class (120 images total) was isolated, leaving approximately 480 images for training. During training, 10% of the training data, constituting around 50 images, was allocated for validation. The DenseNet121 model was employed for training, leveraging its powerful visual feature extraction capabilities for effective resume classification.





Evaluation Metrics: Accuracy: 100% , **Loss**: 0.0059

Classes	F1-Score	Precision	Recall
Resume	1.00	1.00	1.00
Non-Resume	1.00	1.00	1.00

In the context of resume classification:

False Negatives (FN): If the model misclassifies a resume as "Not_Resume," it could lead to overlooking potential candidates, impacting the hiring process.

False Positives (FP):If a non-resume document is misclassified as a "Resume," the consequences are relatively minor in this scenario, as it might not significantly affect the application process.