**RVL-CDIP: Transfer Learning Based Document Classification**

**Mihir Inamdar Param Chordiya**

1. P.I.C.T. - mihirsinamdar@gmail.com
2. P.I.C.T. - param1111.chordiya@gmail.com

.

**Abstract**

This paper presents a method for classifying and searching document images based on deep convolutional neural networks (CNNs) using transfer learning. Deep neural networks can learn hierarchical chains of abstractions between pixels and concise, descriptive representations of scenes or objects. From a document analysis perspective, some popular alternatives are compared with this feature. Using CNNs, we found that (i) features are robust to compression, (ii) CNNs trained on non-documentary images can be used for document analysis tasks, and (iii) regions can be redefined given sufficient training data. Indicates that no learning of unique features is required. Available. A total of 16000 images were used for training, with 1000 images belonging to each class. Below are examples of each class. A dataset from RVL-CDIP was used for data collection.

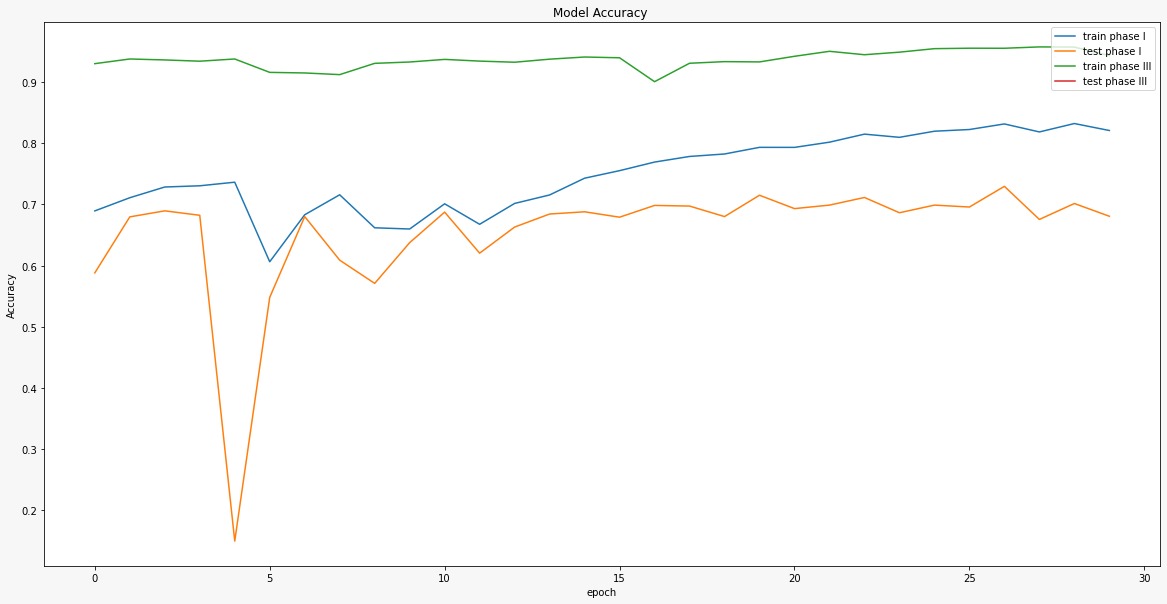
**Feature Selection/Method Description**

The 2D CNN based model includes additional custom dropout and flattened layers created with InceptionResNetV2 and pretrained image mesh weights.

**Experimental Results**

We have trained a total of 90 epochs, achieving an accuracy score of 90.60% with a loss of 0.2907.

**Figures and Tables**

****

Training and testing the InceptionResnetV2 model in three phases..

**Novelty**

InceptionResNetV2 contains two optimal model architectures, ResNet and InceptionNET. The model uses the residual hop vectors from the previous layer and passes them on to the next block in the initial mesh model. This model uses 3x3 convolution to reduce computation time and increase computation speed. You can also convert nxn decompositions to 1xn and nx1 decompositions. To solve the display bottleneck problem, Inceptionnetv2 expanded the feature bank of modules instead of hardening them. As a result, you can prevent the loss of information that occurs when digging. So our approach relies on Inceptionresnetv2.

**Conclusion**

An existing model developed for this dataset did not perform as expected, so we used a model with pre-trained weights.

**Citation Style and Reference Style**

[1] A. W. Harley, A. Ufkes, K. G. Derpanis, "Evaluation of Deep Convolutional Nets for Document Image Classification and Retrieval," in ICDAR, 2015

[2] Zifeng Wu, Chunhua Shen, Anton van den Hengel,

Wider or Deeper: Revisiting the ResNet Model for Visual Recognition, Pattern Recognition, Volume 90, 2019

https://doi.org/10.1016/j.patcog.2019.01.006.

**References**

Ieeexplore.ieee.org. 2022. Analysis of Convolutional Neural Networks for Document Image Classification. [online] Available at: <https://ieeexplore.ieee.org/abstract/document/8270002>

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv. <https://doi.org/10.48550/arXiv.1602.07261>

[Chris Tensmeyer](https://ieeexplore.ieee.org/author/37086295035); [Tony Martinez](https://ieeexplore.ieee.org/author/37282722100) Analysis of Convolutional Neural Networks for Document Image Classification. [online] Available at: https://ieeexplore.ieee.org/abstract/document/8270002