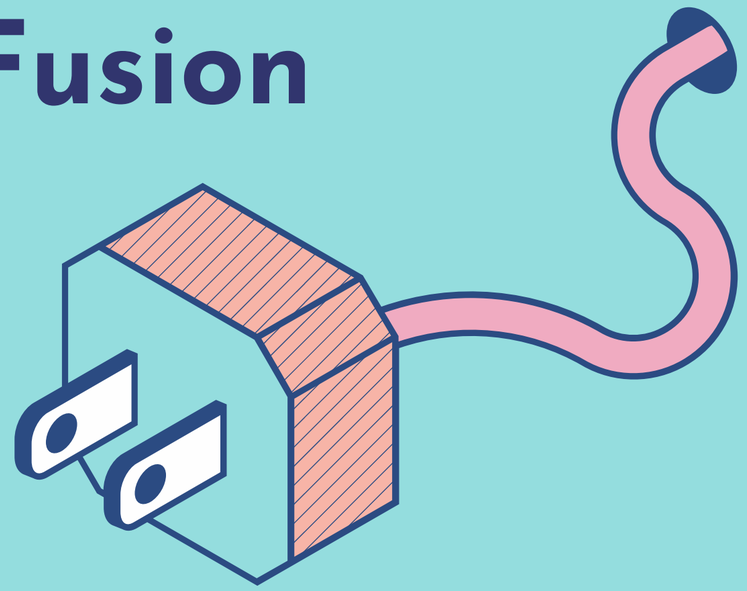
Multimodal AI For Image and Text Fusion

By:
Karan Salunkhe (21070149016)
Shikhar S Dhoke (21070149021)
Shubham Londhe (21070149022)

Multi Modality and Fusion



Dataset-

UPMC Food101

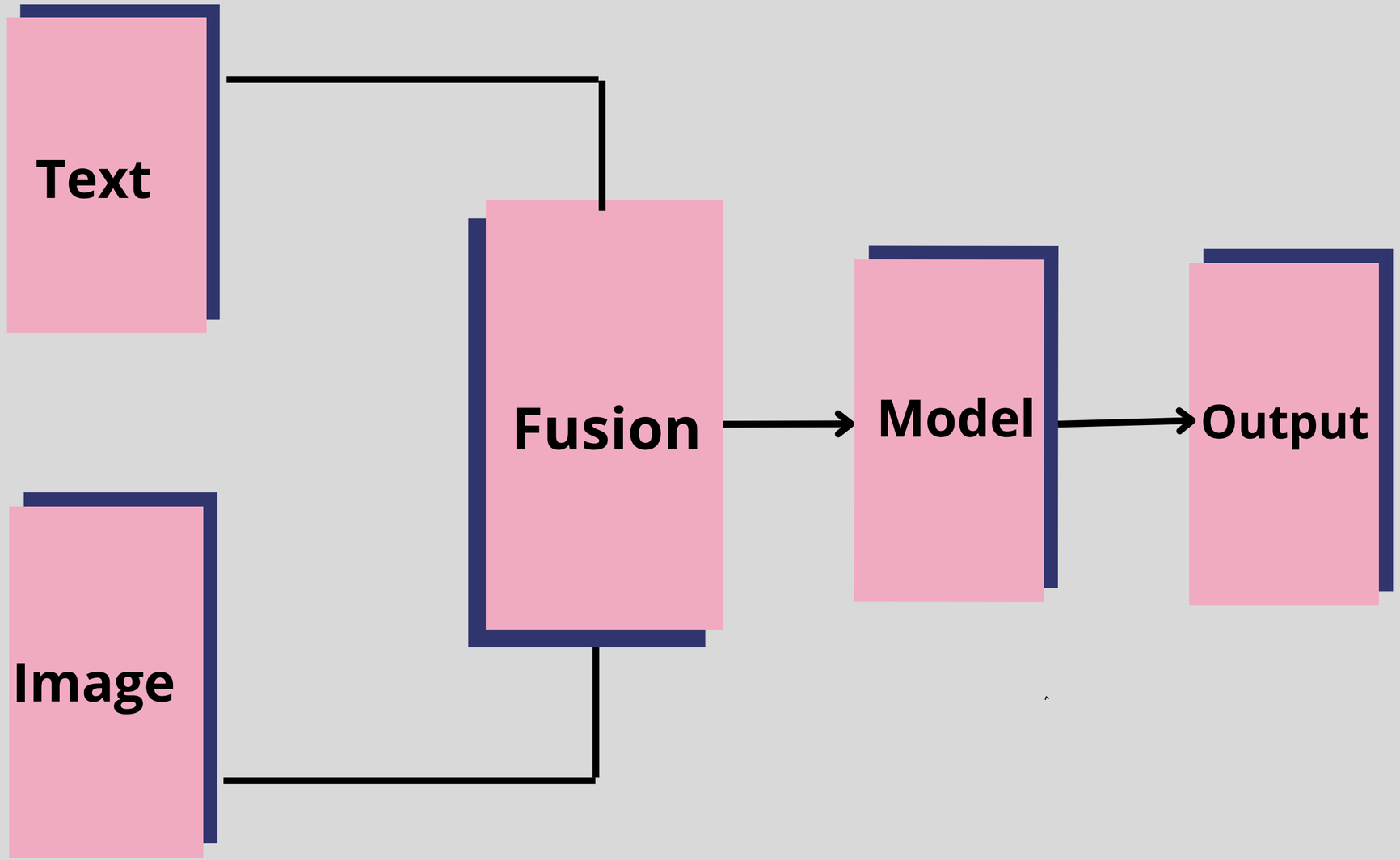
Image

Dataset is composed of a training set containing 67,988 samples and a test set containing 22,716 samples, for a total of 90,704 images.

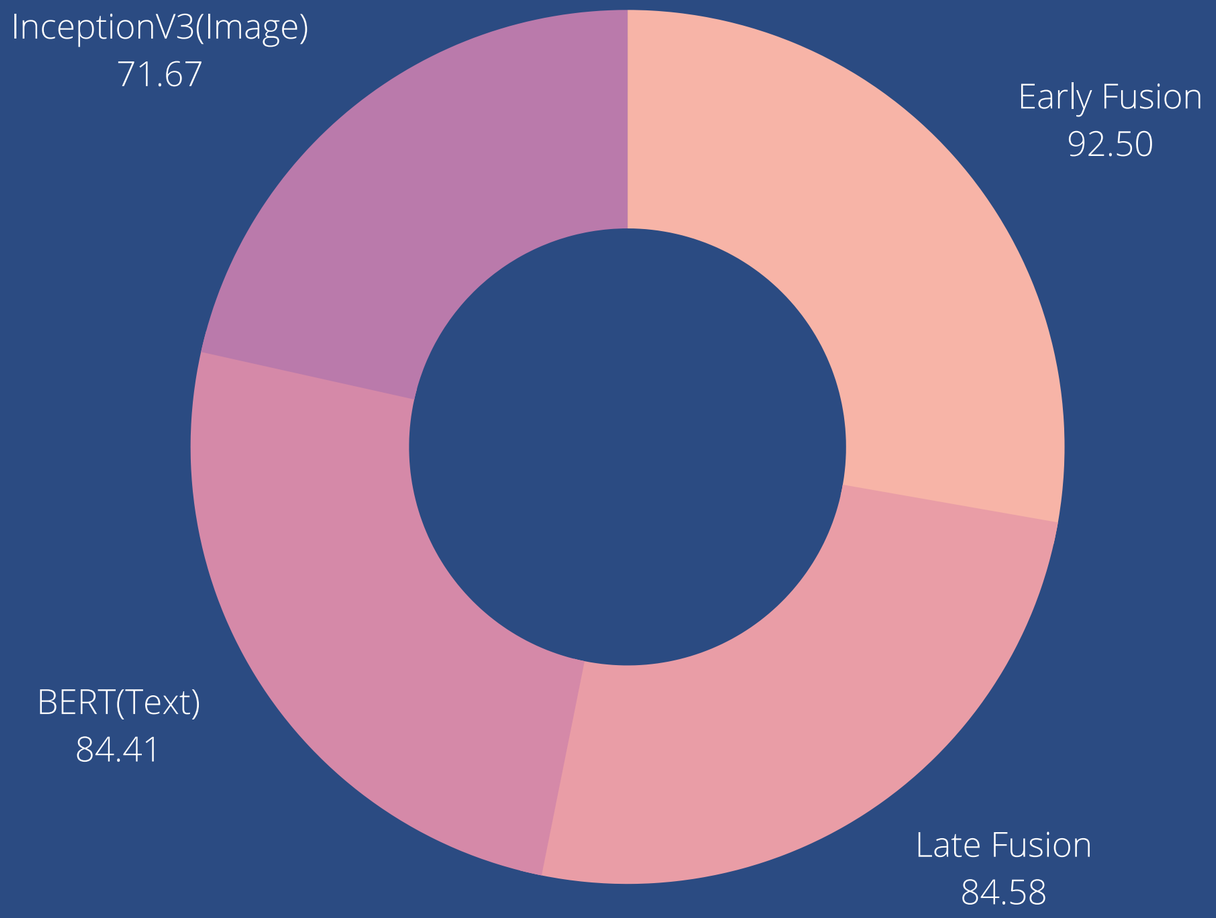
Text

Text documents belonging to 101 classes.

Early Fusion

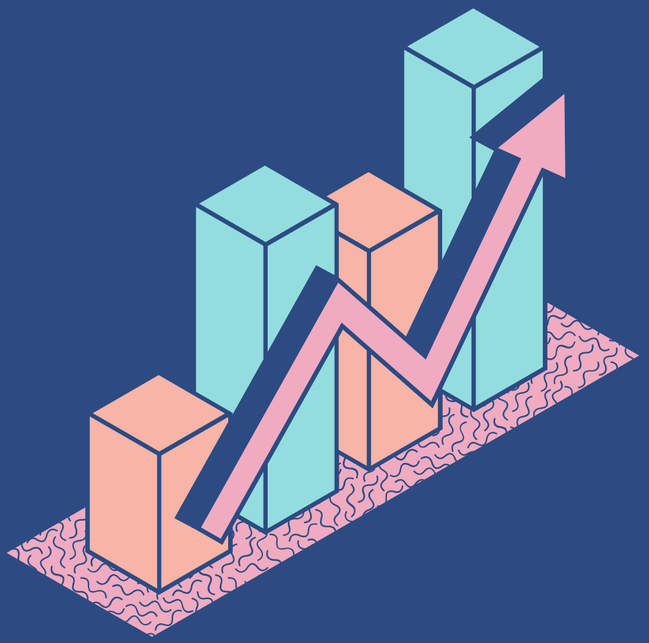


Comparalitive Analysis of Different Models



Results

92.50%



Challenges

1. Large dataset
2. Computation time

References:

1. http://artelab.dista.uninsubria.it/res/research/papers/2020/2020-IVCNZ-Gallo-Food101.pdf
2. https://medium.com/haileleol-tibebu/data-fusion-78e68e65b2d1