Ahmedabad
University

CSE641: Computer Vision: Modern Methods And Applications

Report-5

Group 1

Name	Enrollment No.
Dhyey Patel	AU2240054
Malav Modi	AU2240214
Prem Patel	AU2240010

Introduction

The main objective for this week involved testing modern deep learning models to increase performance in retrieving images from the Flickr dataset. We added MobileNetV3 with Large and Small alternatives together with EfficientNetB3 along with CoAtNet-0 which uses transformer principles to our implementation. The added features had three main purposes which included achieving better accuracy alongside stronger generalization abilities and more effective computational processes.

Implementation of Models

The main objective during this week involved testing complex and advanced models by refining their design infrastructure and training systems to reach optimal performance metrics.

1. EfficientNetB3

Upgrading from EfficientNetB0 to EfficientNetB3 allowed our model to extract improved detailed features. The transition expanded both depth and width of the feature space which delivered better outcomes for tasks involving classification and retrieval.

Received mAP: 0.0569

2. MobileNetV3 (Large & Small)

- The system selected MobileNetV3-Large because it used squeeze-and-excitation with h-swish activation to provide exceptional accuracy alongside speed performance.
- We evaluated MobileNetV3-Small because this version specifically targets mobile and embedded vision applications for assessing lightweight model functionality.
- The attention pooling mechanism was used in both versions to improve spatial feature recognition capabilities.

Received mAP: 0.0544 (MobileNet_V3_small)
Received mAP: 0.0630 (MobileNet_V3_large)

3. CoAtNet-0

The research team integrated CoAtNet-0 as a method to unite the advantages of transformer and convolutional architectural designs.

Received mAP: 0.1018

Optimization Using AttentionPooling2D

We substituted the standard global average pooling operation with AttentionPooling2D to improve spatial region focus by the model. This helped in:

- Highlighting semantically rich features,
- The model becomes more accurate at retrievals through its ability to focus on important regions.
- The method delivers advanced generalization ability with limited extra computational requirements.

Improvement in mobilenet_V2 due to adding this:
Original mAP: 0.0501
Improved mAP: 0.0581

Observations and Results

- All models increased their performance whenever attention pooling techniques were applied.
- The highest accuracy in retrieval rank came from EfficientNetB3 and CoAtNet-0 implementation.
- MobileNetV3-Small provided a high-speed operation combined with accurate results which makes it ideal for mobile platforms and real-time implementations.

Goals for Next Week

- We will execute the following plan during the upcoming week.
- The team must conduct quantitative assessments between all models by measuring mAP as well as Precision@K and Recall@K.
- We will measure the model size with battery requirements and performance speed ahead of deployment.
- We will start implementing person re-identification datasets to evaluate how models perform in particular tasks.