Sean Oldenburger
Dr. Qilin Li
Prof. Ling Li

IMAGE RETRIEVAL:

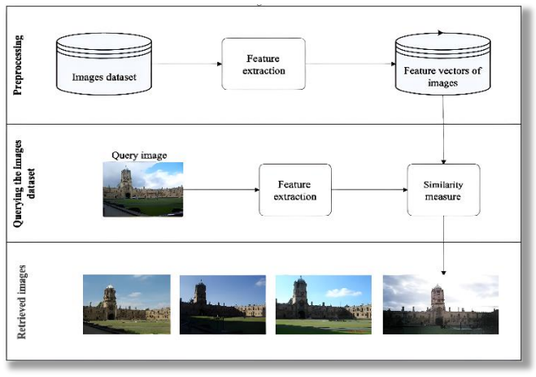
A comprehensive study of representation learning

Curtin University
Pawsey Centre

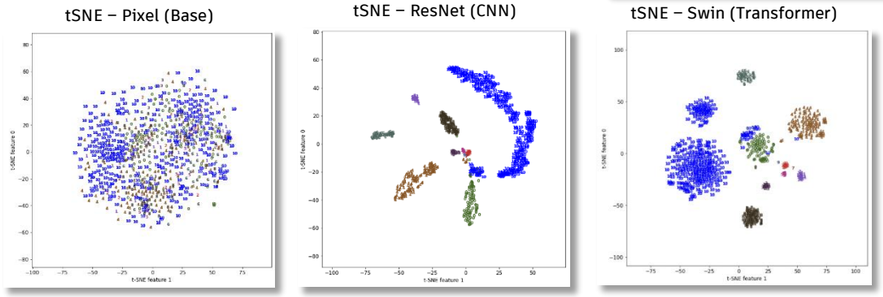
Introduction

My aim of this project was to create a content-based image retrieval system, which returns the top k similar images to a given query based on the features extracted.

These features are extracted through various methods or models, starting from pixel values to SIFT (Scale Invariant Feature Transform) to CNN's (Convolutional Neural Network's) to vision transformers.



Visualization



tSNE (t-distributed Stochastic Neighbor Embedding) is a great way to transform our 1000+ dimension space to 2 so we can visualize the change of a mAP from 0.13 to 0.80 to 0.9.

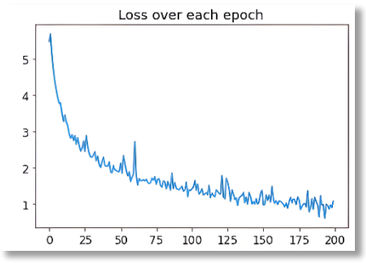
Another great way to visualize data is to look at the direct retrieval output of a query image.

These two visualization methods show how new and different models outperform the last and create a better image retrieval system.

The retrieval output reveals how fine-tuning (ft), compared to the base model, improves the correct images returned.

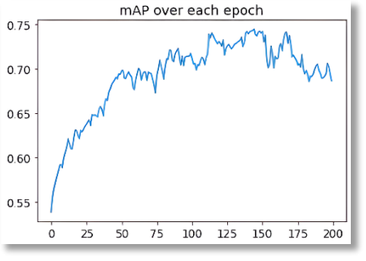
Pixel
▼
ResNet
▼
ResNet ft
▼
Swin
▼
Swin ft

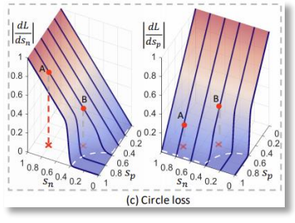
Fine-tuning



One of the most important roles in improving a model's performance is to fine-tune it to a specific task and dataset.

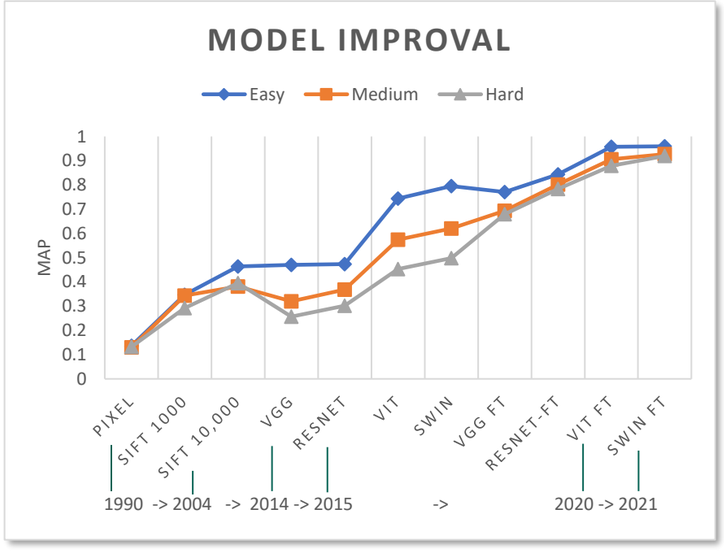
Fine-tuning a model can improve mAP by a significant amount, 0.47 to 0.75!





Using a circle loss function achieves optimal results.

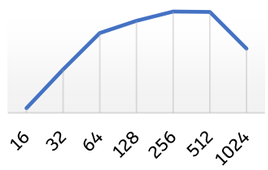
Results



Over the years models improve in design and mAP from 0.46 (2004) to 0.84 (2015) to 0.96 (2021).

The easy oxford dataset increases its accuracy by 83% from a mAP of 0.13 to 0.96.

TESTING PCA:



PCA is a dimensionality reduction technique improving speed and accuracy of a system.

Conclusion

The project was successful, completing all the goals I had set myself. Ending with my own accurate image retrieval system which works for any selected model, using transformers as the recommended.

Transformer models are the most effective for feature extraction, outperforming the well-known CNN's which were SOTA for 6 years.

Image retrieval systems are used commercially in multiple fields and are crucially important to understand and continue to improve.

This study is used in the medical field for retrieving medical plans for similar patient symptoms and widely used in Google image searching.

The authors acknowledge CSIRO, Pawsey Supercomputing, A Backhaus and E Matwiejew for providing resources, training and mentorship during this research project.

References: