Geological Image Retrieval

Present by: Iman Zabett

Geological Image retrieval Approaches

In the area of computer vision and data mining, image retrieval is one of the hot area. We are given an image from the user. We have a large set of images available to us. We want to compute similar images to the given image. There are two main approaches to find similar images:

1- elastic matching of a time series, which is a method widely used in voice recognition and time series comparison.

2- content-based image retrieval (CBIR) tools such as comparison of wavelets, Fourier analysis, or pattern recognition with neural networks.

However, these methods cause a loss of a lot of the image information, or need an extensive training similar to the neural networks.

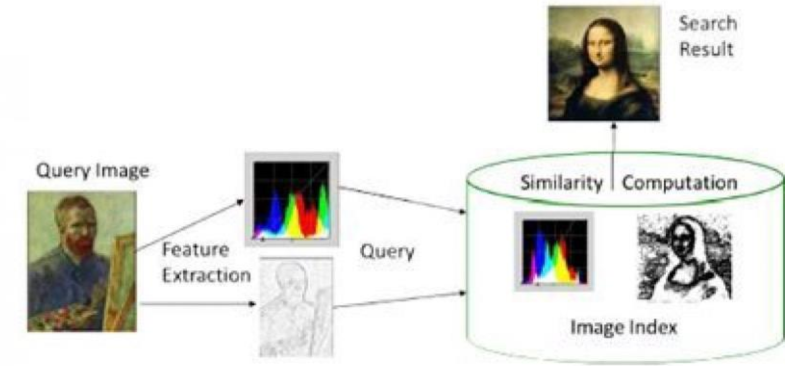
The most used method is the description-based image retrieval using metadata associated with the images, but in an unknown dataset, this method is not effective.

In this work, three methods have been used:

1- Unsupervised approach variational autoencoder (VAE)

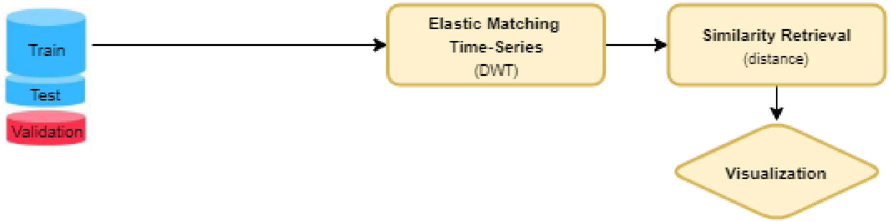
2- Semi-supervised approach Perceptual VAE

3- Supervised approach of Transfer Learning with VGG network

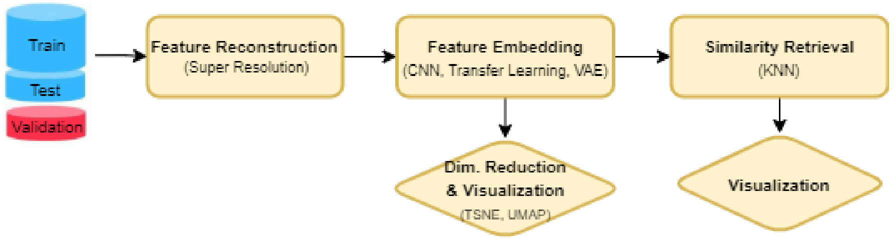


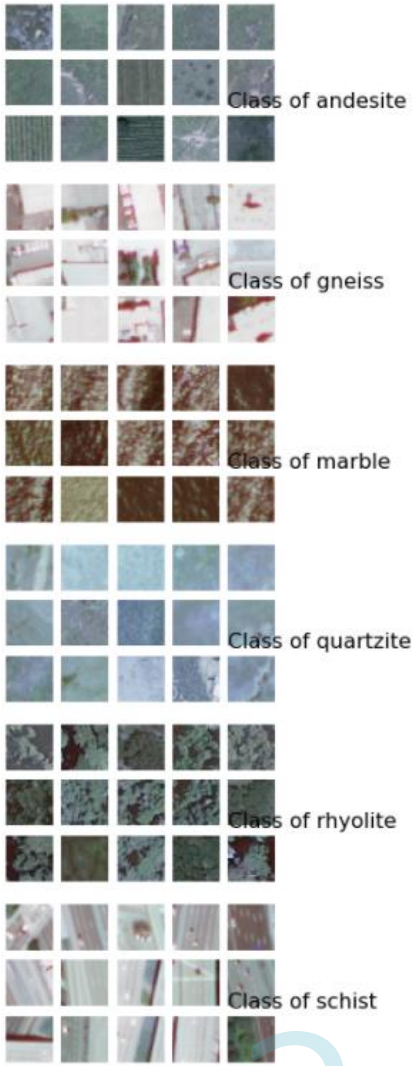
Content-Based image retrieval schema

Schema of the geological elastic search image retrieval pipeline

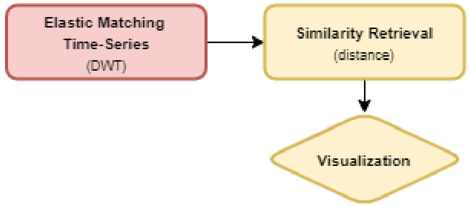


Schema of the geological content-based image retrieval (CBIR) pipeline





Overview of Geo dataset with six classes



Dynamic time warping (DTW)

[Notebook DTW]

Dynamic time warping (DTW)* is an elastic matching algorithm used in pattern recognition. DTW finds the optimal warp path between two time series. DTW is used as a distance metric, often implemented in speech recognition, data mining, robotics, and in this case image similarity

There are some advantages and disadvantages for this:

- No need to use of any metadata
- No need to use concept-based image indexing
- Time consuming
- not reproducible

Two methods for extracting data from image have been used. Quick method which is only extract slice of 3-CH matrix which dramatically reduced runtime of search through geo dataset (~ 13 min). And Exhaustive method which uses entire data (3-CH matrix) of images, which requires high amount of time to search through entire geo dataset (~ 1 hour). The images show the comparisons between 20 most similar images regarding to two methods and below shows the comparison between their 2-way plots for two sample image w.r.t. the query image (first image of the row)

Other image Similarity Metrics using in this work:

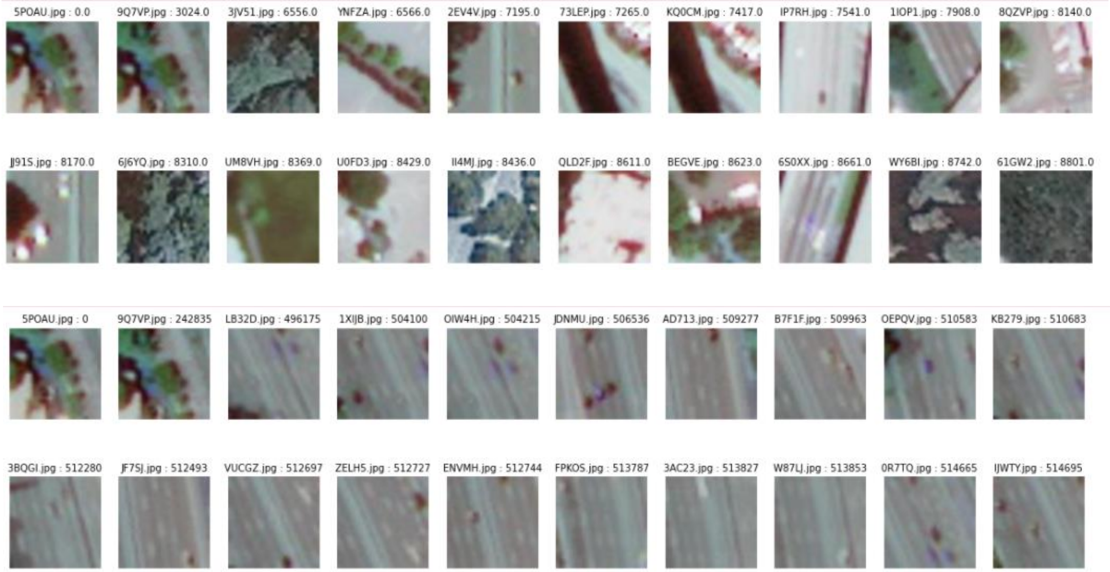
- Peak Signal to Noise Ratio (**PSNR**)

- Structural SIMilarity method (**SSIM**)

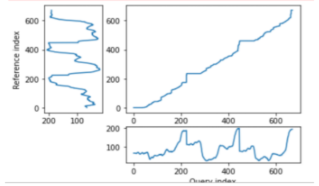
*Reference: Lei, Hansheng, and Venu Govindaraju. "Direct image matching by dynamic warping." In *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pp. 76-76. IEEE, 2004.

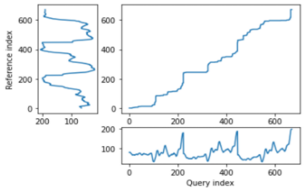
Quick DTW

Exhaustive DTW

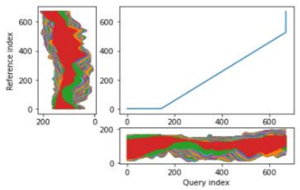


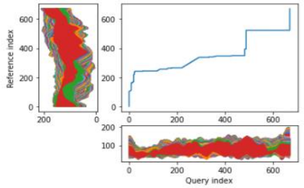
Quick DTW
9Q7VP and
3JV51

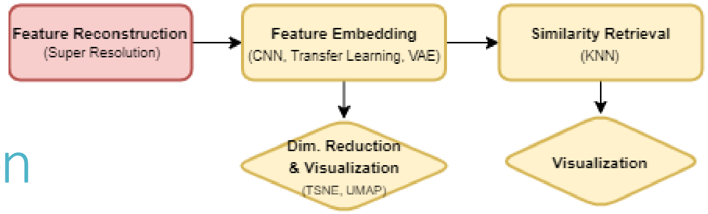




Exhaustive DTW
9Q7VP and
3JV51







Super Resolution

Deep Back-Projection Networks (DBPN) for Single Image Super-resolution

Motivation:

We have very low information of our experimental dataset.

The proposed dataset contains very low resolution images representing six different geological patterns.

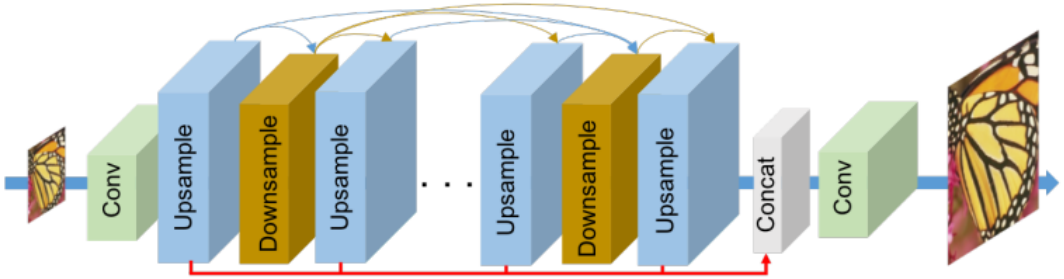
For example we don’t know how big the actual images, or the images are tiles of a bigger image. If they have been down-sampled, how many times they have lower resolution than their actual size.

Or even we don’t know how close they have been photographed (are they satellite images or close-up images)

DBPN:
Consists of three main phases:

- Initial feature extraction with Conv layer,
- Back-projection with multiple dense up-/down- projection units to form feature maps (HR and LR)
    - scale up/down, scale down/up, clac residual, scale residual up/down, output feat map
- Image reconstruction by concatenating feature maps produced in projection units and a conv layer

Reference: Haris, Muhammad, Greg Shakhnarovich, and Norimichi Ukita. "Deep back-projection networks for single image super-resolution." *arXiv preprint arXiv:1904.05677* (2019).



(d) Iterative up- and down-sampling

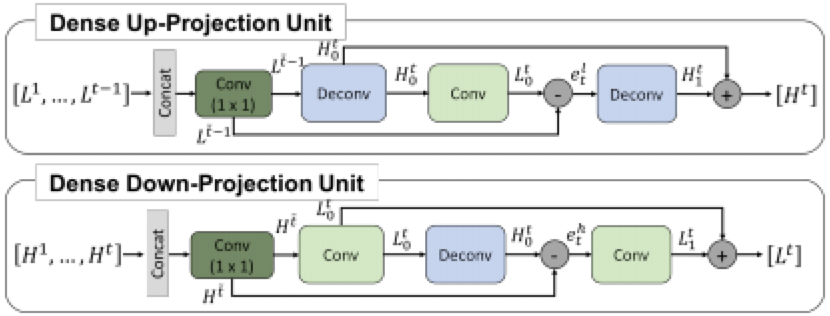


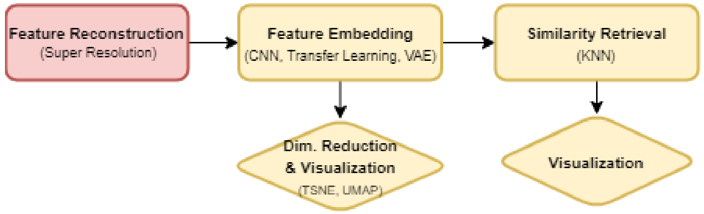
TABLE 8
Quantitative evaluation of state-of-the-art SR algorithms: average PSNR/SSIM for scale factors 2×, 4×, and 8×. Red indicates the best and blue indicates the second best performance.

Method	Scale	# Parameters (M)	Set5 PSNR	Set5 SSIM	Set14 PSNR	Set14 SSIM	BSDS100 PSNR	BSDS100 SSIM	Urban100 PSNR	Urban100 SSIM	Manga109 PSNR	Manga109 SSIM
Bicubic	8	-	24.39	0.657	23.19	0.568	23.67	0.547	20.74	0.516	21.47	0.647
A+ [23]	8	-	25.52	0.692	23.98	0.597	24.20	0.568	21.37	0.545	22.39	0.680
SRCNN [17]	8	0.05M	25.33	0.689	23.85	0.593	24.13	0.565	21.29	0.543	22.37	0.682
FSRCNN [18]	8	0.01M	25.41	0.682	23.93	0.592	24.21	0.567	21.32	0.537	22.39	0.672
VDSR [12]	8	0.66M	25.72	0.711	24.21	0.609	24.37	0.576	21.54	0.560	22.83	0.707
LapSRN [13]	8	0.81M	26.14	0.738	24.44	0.623	24.54	0.586	21.81	0.582	23.39	0.735
MS-LapSRN [52]	8	0.22M	26.34	0.753	24.57	0.629	24.65	0.592	22.06	0.598	23.90	0.759
MSRN [53]	8	6.3M	26.59	0.725	24.88	0.5961	24.70	0.541	22.37	0.598	24.28	0.752
D-DBPN [31]	8	23.1M	27.25	0.785	25.14	0.649	24.91	0.602	22.72	0.630	25.14	0.798
EDSR [30]	8	43.2M	26.97	0.775	24.94	0.640	24.80	0.596	22.47	0.620	24.58	0.778
RCAN [15]	8	16M	27.31	0.787	25.23	0.651	24.98	0.606	23.00	0.645	25.24	0.803
SAN [54]	8	15.7M	27.22	0.783	25.14	0.648	24.88	0.601	22.70	0.631	24.85	0.791
DBPN-RES-MR64-3	8	23.1M	27.51	0.793	25.41	0.657	25.05	0.607	23.20	0.652	25.71	0.813

Super Resolution (Contd)

[Notebook link]

Comparison between DBPN and Bicubic approach to reconstruct high resolution image

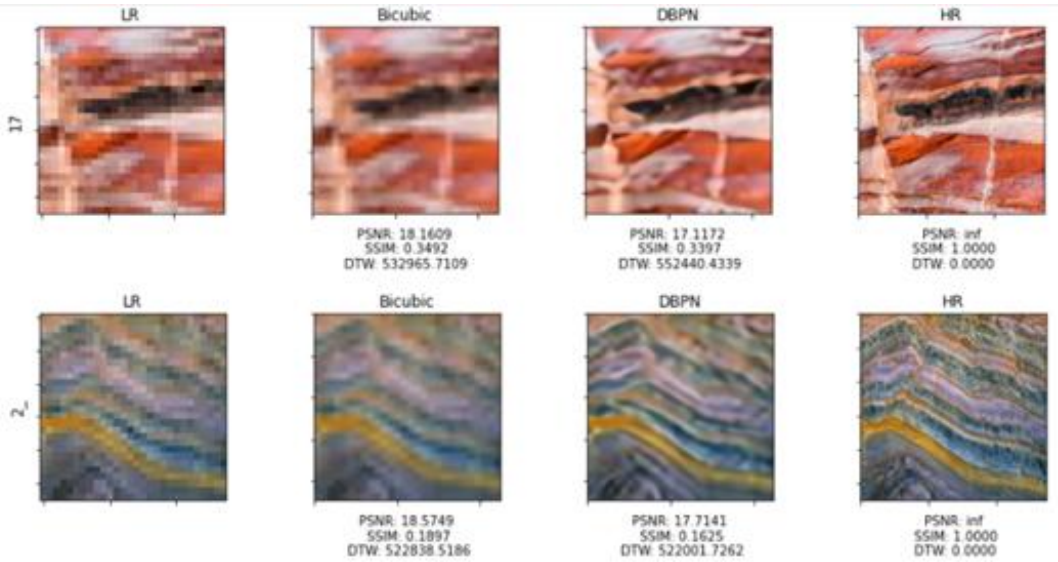


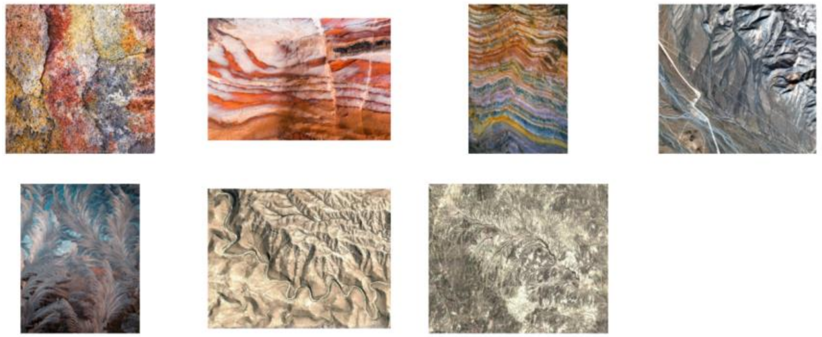
Motivation:

The DBPN model has been trained with variety of high resolution images (DIV2K). Then the pretrained model used to

A comparison has been made over cropped images from high-resolution geological images

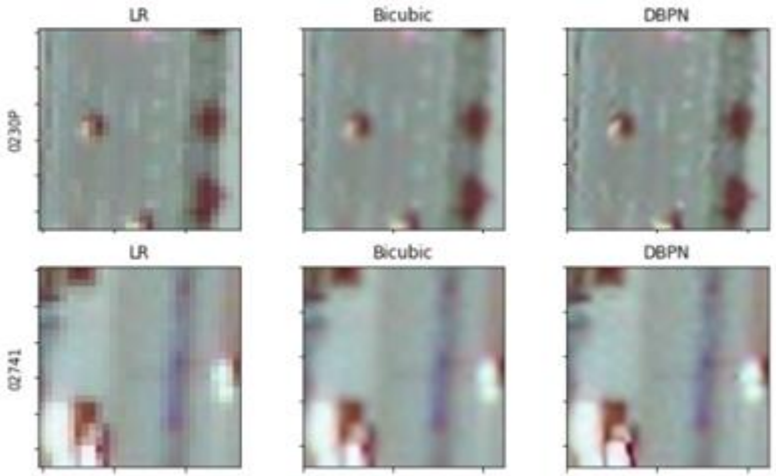
- Cropping from HQ images to 224*224 make the HR images (4th column)
- Down-scaling the HR cropped fo 24*24 and make LR images (1st column)
- Up-scaling the first column with Bicubic and make the 2nd column
- Up-scaling the first column using DBPN model and make the 3rd column





Preview of HQ geological images

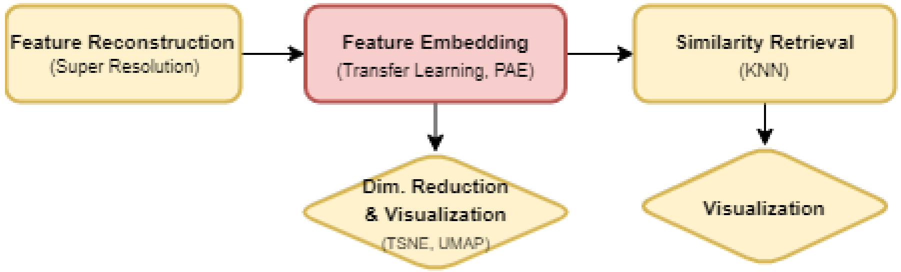
Crop from the HQ images to test model



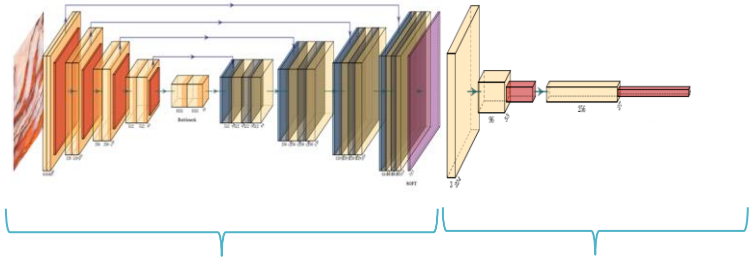
Apply DBPN on our Geo dataset samples

Perceptual Autoencoder

[Notebook link]



perceptual loss recently proposed as an alternative to element-wise loss to improve autoencoder embeddings for downstream prediction tasks.

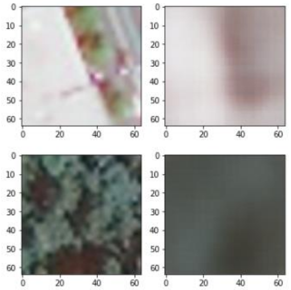


Variational Autoencoder

Perceptual Network
Alexnet
(2 feature layers)

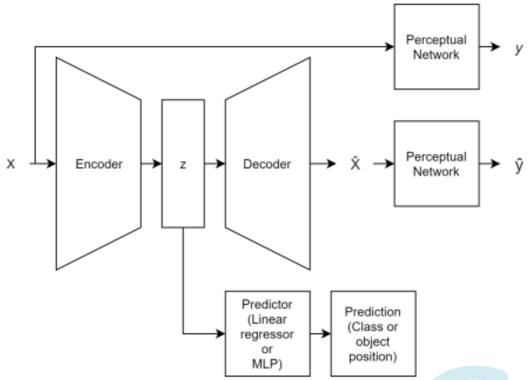
In this approach, the perceptual network has been developed as follows:

- an Alexnet has been trained with train Geo dataset with up to %97.50 accuracy
- Then only the first two layers of feature layers of the Alexnet was used
- A sigmoid function has been added after the last RELU layer



Sample of input and reconstructed image of VAE

Then the perceptual network was used follow with the variational autoencoder, to get the constructed images of the VAE as an input. This forms the state-of-the-art P-VAE network that can be trained against the training geo dataset.

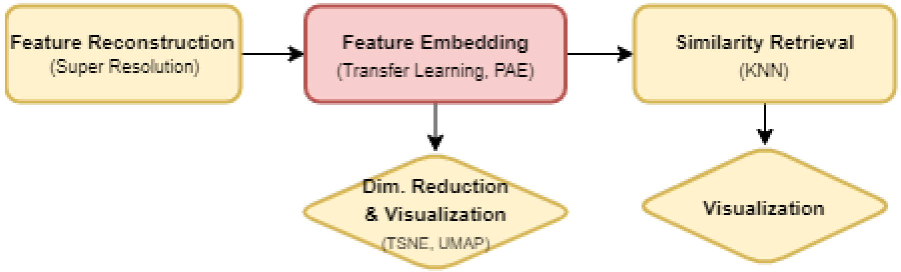


Schema of Perceptual VAE network

Reference: Pihlgren, Gustav Grund, Fredrik Sandin, and Marcus Liwicki. "Improving image autoencoder embeddings with perceptual loss." In 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1-7. IEEE, 2020.

Transfer Learning
using VGG19bn

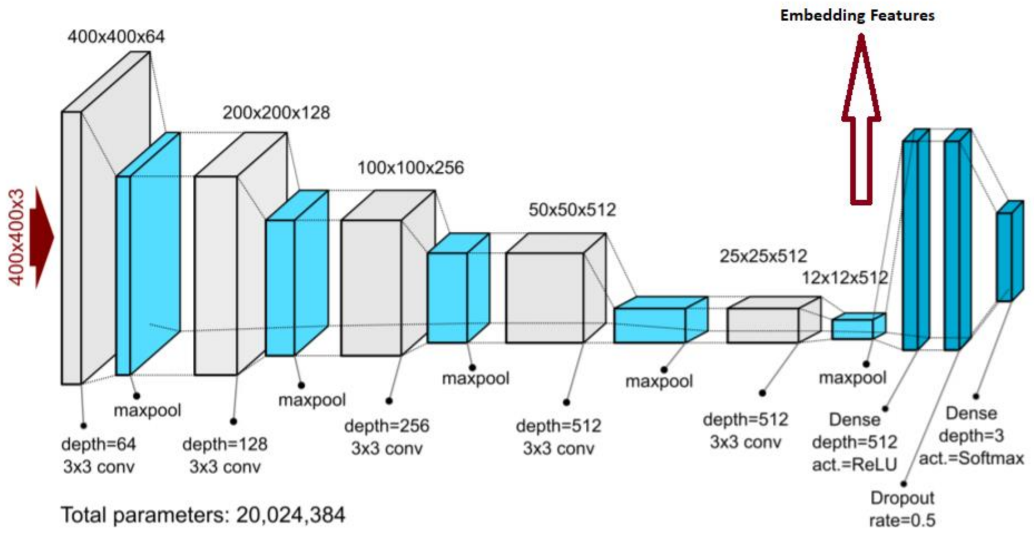
Notebook link



The VGG-19 CNN architecture is reported to achieve high accuracies for image processing large datasets such as ImageNet. VGG-19 model has roughly 143 million parameters. The VGG-19 contains 19 trainable layers including convolutional and fully connected layers as well as max pooling, and dropout.

Batch normalization (also known as batch norm) is a method used to make artificial neural networks faster and more stable through normalization of the input layer by re-centering and re-scaling.

The training over the geo dataset after 50 epochs, achieved up to %99 accuracy over the test dataset. The VGG19bn has been modified to classify for 6 classes. After training, the features (after the last feature layers) will be extracted for validation dataset in order to retrieve the most similar images



Architecture of VGG19bn

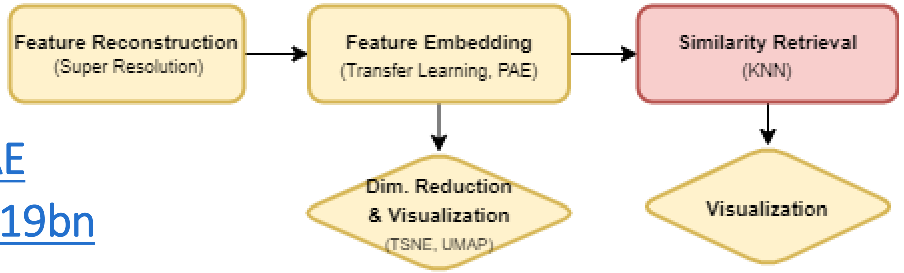
Similarity Retrieval with KNN

comparisons between VAE, P-VAE, and VGG19bn

Notebook VAE

Notebook P-VAE

Notebook VGG19bn



The comparison is being made using an image query.

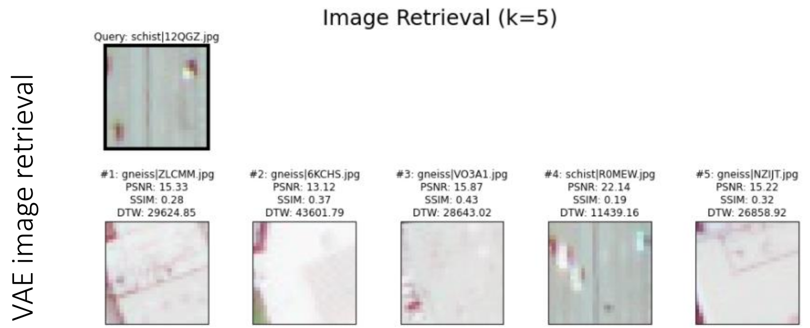
Then K-Nearest Neighbor (KNN) is being use to find the closest images in the feature spaces.

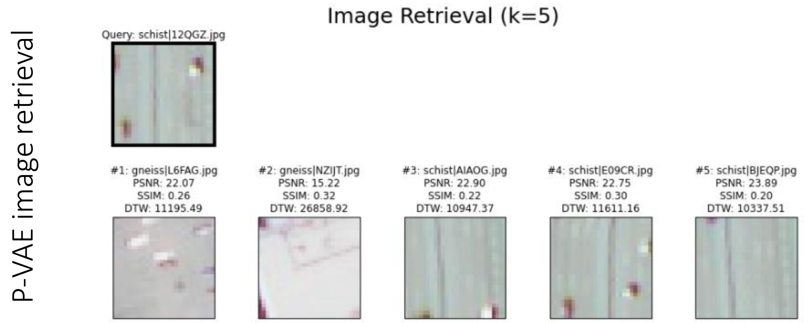
The feature space for VAE (128 dimensions), P-VAE (128 dimensions), and VGG19bn (2048 dimensions).

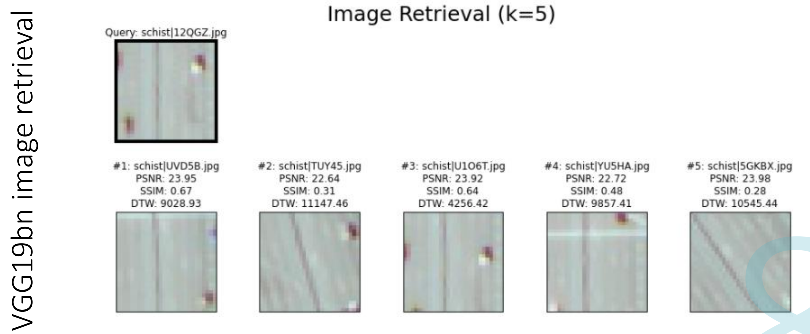
The entire train geo set is used to fit the KNN with "cosine" metric to calculate the distance of samples.

There are three similarity metrics (PSNR, SSIM, DTW) being used to calculate similarity of each retrieved image (nearest image) with the query image.

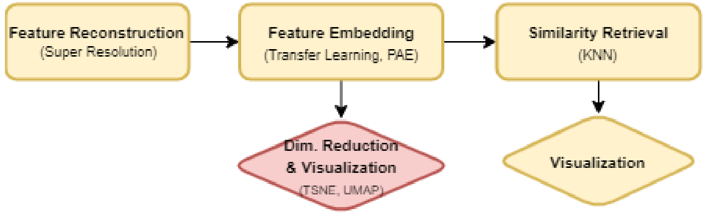
The result shows that VGG19bn has better overall performance over P-VAE and P-VAE has better overall performance over VAE.







Dimensionality Reduction for Visualization



t-Distributed Stochastic Neighbor Embedding (t-SNE)

TSNE takes much longer and is designed to retain local structure within the data.

The transformation has done for entire data (trian + test) features, since there is no "transform" function to transform test features after building TSNE embedding model. The visualization however is only for test features in 2D plane.

Uniform Manifold Approximation and Projection (UMAP)*

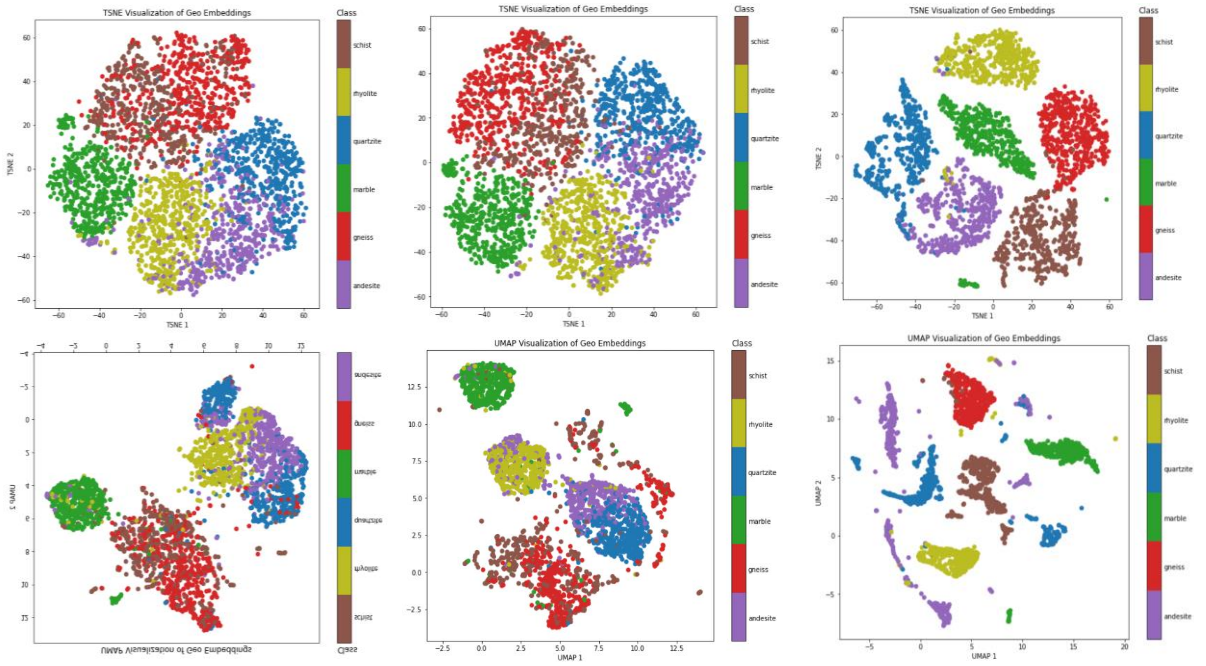
UMAP is generally quicker and is designed for a balance between local density and global/topological structure in the embedding. Also "transform" is available.

The algorithm is founded on three assumptions about the data:

1. The data is uniformly distributed on a Riemannian manifold;
2. The Riemannian metric is locally constant (or can be approximated as such);
3. The manifold is locally connected.

Results

The results show that UMAP had better performance to maintain global structure. Also VGG19bn (transfer learning) outperforms over VAE and P-VAE in order to have distinct clusters of classes with enough distance from other classes



TSNE/UMAP for VAE TSNE/UMAP for P-VAE TSNE/UMAP for VGG19bn

* Reference: Parra-Hernández, Ronald M., Jorge I. Posada-Quintero, Orlando Acevedo-Charry, and Hugo F. Posada-Quintero. "Uniform Manifold Approximation and Projection for Clustering Taxa through Vocalizations in a Neotropical Passerine (Rough-Legged Tyrannulet, Phyllomyias burmeisteri)." Animals 10, no. 8 (2020): 1406.

Thank you!