Where's Waldo?
Instance-level Image Retrieval

Mikayla Pugel, David Knapp, Ferdous Alam

Berkeley
UNIVERSITY OF CALIFORNIA

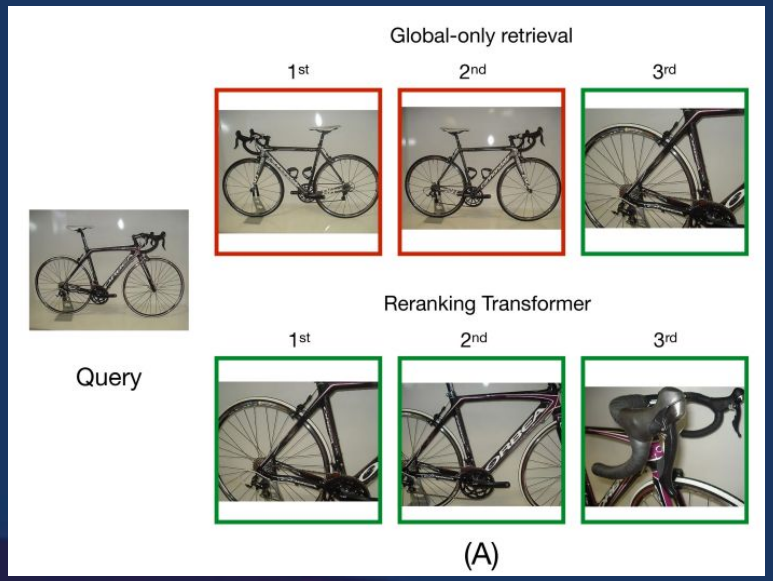
Agenda

1. Introduction/Project Objective
2. Impact of Solution
3. Technical Approach
4. Evaluation
5. Demo of Solution
6. Testing Another Dataset
7. Summary

Berkeley
UNIVERSITY OF CALIFORNIA

Instance-level Image Retrieval

The task of searching in a large database for images that match an object in a query image.



Berkeley
UNIVERSITY OF CALIFORNIA

Project Objective

Primary goal : Improve the process of the combined feature extraction and reranking transformers instance level image retrieval through transformations of the images and optimizing the model's variables.

Secondary goal: Test to see how the changes we have made impact the retrieval of different images/datasets

Berkeley
UNIVERSITY OF CALIFORNIA

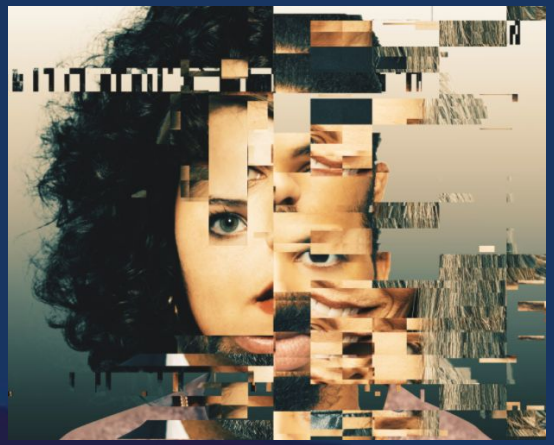
Opportunity Space

Tourism

Artwork

E-commerce / Retail

Law Enforcement



Berkeley
UNIVERSITY OF CALIFORNIA

Data

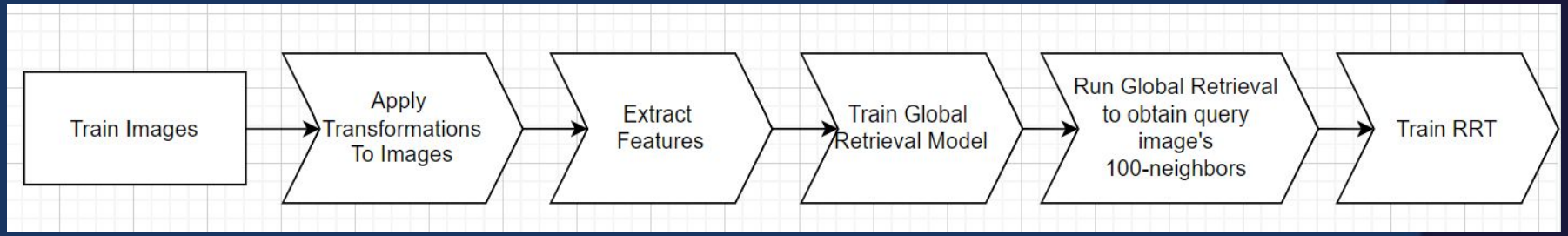
Stanford Online Products (SOP):

- 120,053 images, 59,551 for training, 60,502 for testing, 22,634 classes of 12 super classes
- Super classes include: bicycle, cabinet, chair, coffee maker, fan, kettle, lamp, mug, sofa, stapler, table, toaster

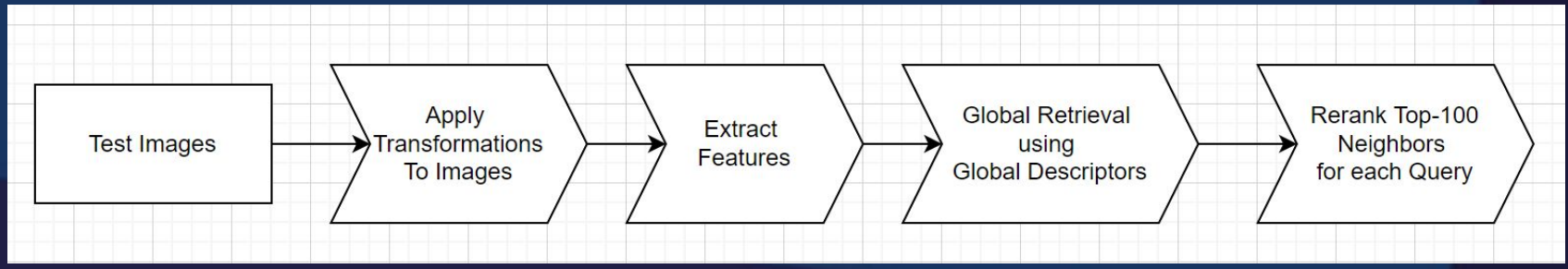
Berkeley
UNIVERSITY OF CALIFORNIA

Process Overview

Training Process:



Testing Process:



Berkeley
UNIVERSITY OF CALIFORNIA

Model Specifications

Image Processing During Training:

- Each image randomly cropped to 224 by 224
- Each image is randomly flipped

Image Processing During Testing:

- Each image is resized to 256 by 256 then cropped at the center to 224 by 224

For Feature Extraction:

- Using ResNet50 features are extracted and which creates 49 (7 by 7) local descriptors for each image
- The global descriptors are obtained by spatially averaging the local responses
- Both the global and local descriptors are linearly projected to a dimension of 128

Model Architecture:

Global Model:

- trained with contrastive loss
- Batch size = 800
- Test batch size = 800
- Only one image scale is used so no scale embedding is used
- trained using SGD with Nesterov momentum for 100 epochs
- learning rate of 0.001
- weight decay of 0.0004
- momentum of 0.9

RRT Model:

- trained on top of the global model
- using both freezing and finetuning CNN backbone
- trained with AdamW for 100 epochs
- learning rate of 0.0001

Berkeley
UNIVERSITY OF CALIFORNIA

Testing Specifications

1. CO:
   - a global retrieval model trained with contrastive loss
   - using metric learning protocol
2. CO + RRT (frozen)
   - the pretrained CO remains frozen and an extra linear layer is used to reduce the dimension of the local descriptors to 128
3. CO + RRT (finetuned)
   - same architecture but the backbone as fine tuned

Berkeley
UNIVERSITY OF CALIFORNIA

Initial experiments with transformations on test images

Method	R@1	R@10	R@100
Baseline (CO)	*80.7*	*91.9*	*96.9*
Baseline with Reranking (frozen)	*81.8*	*92.4*	*96.9*
Baseline with Reranking (finetuned)	*84.5*	*93.2*	*96.6*
Reranking with RandomPerspective()	69.3	87.8	96.6
Reranking with grayscale transform	67.2	87.7	96.6
Reranking with contrast adjustment	71.0	88.2	96.6

Berkeley
UNIVERSITY OF CALIFORNIA

Experiments Continued

Method	R@1	R@10	R@100
Baseline with RRT (finetuned)	*84.5*	*93.2*	*96.6*
Reranking removing cropping and resizing	86.3	94.2	97.1
Reranking removing random horizontal flip	84.44	93.25	96.7
Reranking with Gaussian Blur	84.4	93.21	96.7
Reranking with Normalization	12.83	32.33	65.97

Berkeley
UNIVERSITY OF CALIFORNIA

Example Results

Image:

NN_1:

NN_2:

NN_3:

Berkeley
UNIVERSITY OF CALIFORNIA

Evaluation by Image Category

Category	Accuracy@1
Bicycle	98.2%
Cabinet	94.4%
Chair	92.0%
Coffee Maker	92.5%
Fan	91.4%
Kettle	94.2%
Lamp	92.5%
Mug	96.8%
Sofa	93.7%
Stapler	95.4%
Table	91.0%
Toaster	92.0%

Berkeley
UNIVERSITY OF CALIFORNIA

Demo

Testing Another Dataset

Products-10k:

- 150,000 photos and 10,000 different products
- product categories include Fashion, food, healthcare, and household commodities

Results:

	R@1	R@10	R@100
10K Products Dataset	**34.73**	**54.07**	**73.42**

Berkeley
UNIVERSITY OF CALIFORNIA

Summary

- Tested new image transformations to improve the current image recognition process results.
- Created prototype process that allows users to use a new photo, crop the photo in any way they like, and see the top retrieved images for that new image.
- Tested another products dataset to see if the results transcended datasets.

Berkeley
UNIVERSITY OF CALIFORNIA

Thank You!

Any Questions?

Berkeley
UNIVERSITY OF CALIFORNIA

References

1. https://www.adobe.com/creativecloud/photography/discover/gaussian-blur.html
2. https://cvexplained.wordpress.com/2020/07/20/10-1-what-are-image-descriptors-feature-descriptors-and-feature-vectors/
3. https://towardsdatascience.com/instance-level-recognition-6afa229e2151
4. https://filebox.ece.vt.edu/~jbhuang/teaching/ece6554/sp17/lectures/Lecture_02_Instance_Recognition.pdf
5. https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939
6. https://paperswithcode.com/method/adamw
7. https://towardsdatascience.com/stochastic-gradient-descent-clearly-explained-53d239905d31
8. https://towardsdatascience.com/the-beginners-glossary-of-neural-network-terms-a9617354078
9. https://inside-machinelearning.com/en/why-and-how-to-normalize-data-object-detection-on-image-in-pytorch-part-1/
10. https://www.kaggle.com/c/products-10k
11. https://ai.googleblog.com/2020/09/advancing-instance-level-recognition.html

Berkeley
UNIVERSITY OF CALIFORNIA

Terms Defined

Finetuning – Implements transfer learning in that the model will not only train its current stage, but finetune all the previous stages

Frozen – Freezes all previous stages during the training of the current stage

Weight Decay: regularization technique by adding a small penalty to the loss function, this is normally used to prevent overfitting and avoids exploding gradient

Learning Rate: controls how much we are adjusting the weights of our network with respect to the loss gradient. The lower the value, the slower we travel along the downward slope

Momentum: accelerates gradient descent in the relevant direction and dampens oscillations

Nesterov momentum: An extension of momentum that involves calculating the the decaying moving average of the gradients of projects positions in the search space rather than the actual positions themselves. This allows the search to slow while approaching the minima.

Contrastive Loss: takes the output of the network for a positive example and calculates its distance to an example of the same class and contrasts that with the distance to negative example.

Berkeley
UNIVERSITY OF CALIFORNIA

CNN Explained

- the first layer in a CNN is always a convolutional layer
- next CNN contains pooling layers which simplify the data by reducing its dimensionality (shortening training time and helps curb the problem of overfitting)
- last layers are the fully connected layers, in which every neuron in the first is connected to every neuron in the next. This process looks at which features most accurately describe the specific classes which results in a simple vector of probabilities

AdamW

- AdamW is a stochastic optimization method that modifies the typical implementation of weight decay in Adam by decoupling weight decay from the gradient update.

Why AdamW matters. Adaptive optimizers like Adam have… | by Fabio M. Graetz | Towards Data Science

Stochastic Gradient Descent

- samples are selected randomly instead of the whole data set for each iteration
- batch denotes the total number of samples from a dataset that is used for calculating the gradient for each iteration
- since only one sample from the dataset is chosen at random for each iteration, the path taken by the algorithm to reach the minima is usually noisier than normal GD
- it may also require more iterations than normal GD, but computationally it is much less expensive

Example Results

Image:

NN_1:

NN_2:

NN_3:

Berkeley
UNIVERSITY OF CALIFORNIA