Supplementary Material for Vis2Rec: A Large-Scale Visual Dataset for Visit Recommendation

Michaël Soumm[1] michael.soumm@cea.fr
Adrian Popescu[1] adrian.popescu@cea.fr
Bertand Delezoide[2] bertrand.delezoide@amanda.com

[1]Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France
[2]Amanda, 34 Avenue Des Champs Elysées, F-75008, Paris, France

Figure 1: Samples of collected images grouped by user

1. Dataset description

We describe here a few additional statistics that we believe could be useful for Vis2Rec understanding.

1.1. Collected images

Metadata. Each of the 7,158,454 collected images comes with its set of metadata, including notably

- a time-stamp corresponding to the `datetaken` attribute of the Flickr API, which could be useful for time-aware recommender systems.
- image title and user tags for text processing purposes
- GPS coordinates for half of the images

Images. A random sample of user images is present on figure 1. User images are varied and do not always depict POIs.

1.2. Visual matching

Training details. We used the DELG model in inference using the official implemtentation[1], which we slightly modified for optimization. In particular, we first generate 2048-dimentionnal global features for all collected of Vis2Rec and all images in GLv2. We then use a *k-nn* algorithm to find the 20 closest images of GLv2 for each image in Vis2Rec . Finally, for each potential match, we generate up to 1000 128-dimensional local descriptors and perform geometric verification using RANSAC. Since the DELG implementation only supports a batch size of 1 image, we parallelized the inference on 576 Intel Xeon Gold 6154 3.00GHz CPUs.

Matching scores. Figure 2 shows samples of matchings 10 point scores ranges from 10 to 60. As we see qualitatively, matches with a score lower than 40 sometime show failures. An analysis on 500 images in each range showed that matches of a score higher than 40 presented a false pos-

[1]https://github.com/tensorflow/models/blob/master/research/delf/delf/python/delg/DELG_INSTRUCTIONS.md

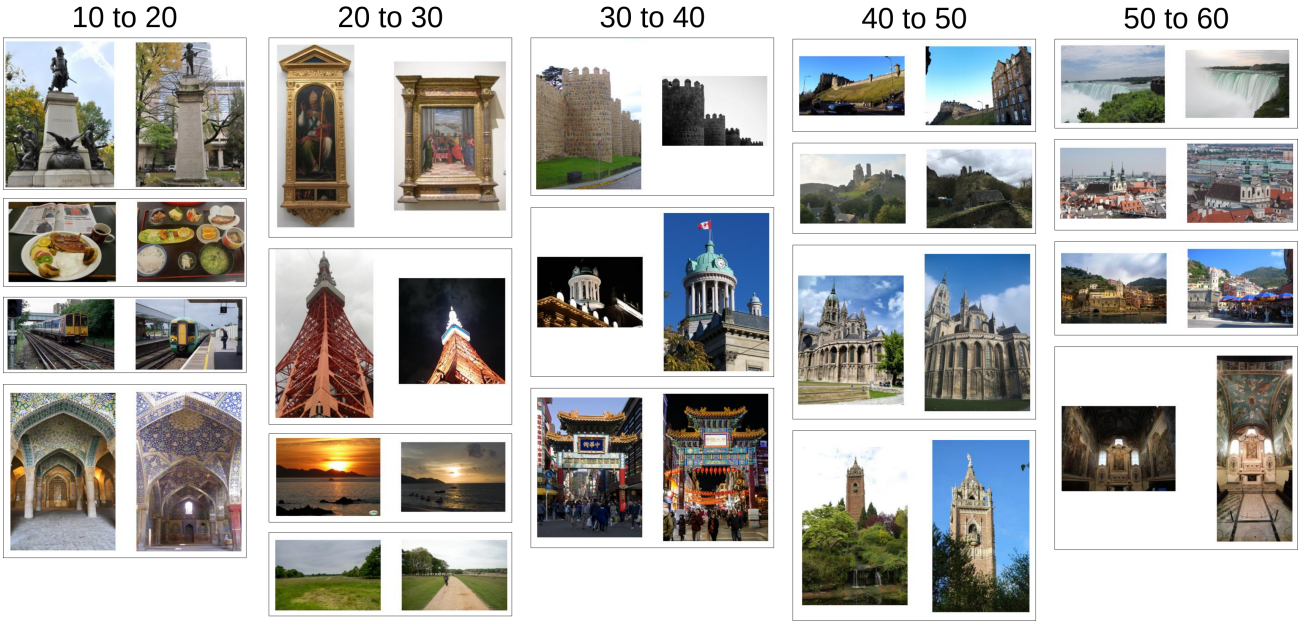

Figure 2: Samples of matched pairs grouped by matching scores

itive rate less than 2%.

Geographic post-processing. We use a two step post processing to clean the matches. First, we use a geolocated subset of Vis2Rec to identify parasite images in GLv2 by selecting images which matched with at least 5 images and only with images taken 15kms away from the predicted landmarks. Secondly, we use the predicted landmarks to infer to determine the most sure POI visit for each each and each day called the *anchor* visit. We remove absurd predicted matches for each day more than 100kms away from the *anchor* visit. A verification on the geolocated subset of Vis2Rec shows that this two step processing significantly reduces the number of geographical outlier detections.

Annotation process. Visits selected for the test *target* set with matching scores less than 40 are manually verified by three annotators. An efficient interface (shown in figure 3) was designed to efficiently annotate 10k images, where simply clicking on a pair would validate it. A match was counted in if at least two annotators agreed on its validity, and as a result, only 56% of the matches were counted as valid.

Annotation results. The sub-sample of 5k annotated images provides for enough data to verify our thresholds. Figure 4 shows the number of image pairs correctly and incorrectly matched depending on their matching score. As we can see, half of the images were mismatched mostly because they have a low matching score, which corresponds to images not depicting POIs or matches perturbed by foreign objects. Looking at the precision obtained at each matching score (Figure 10), we see that a matching score of at least

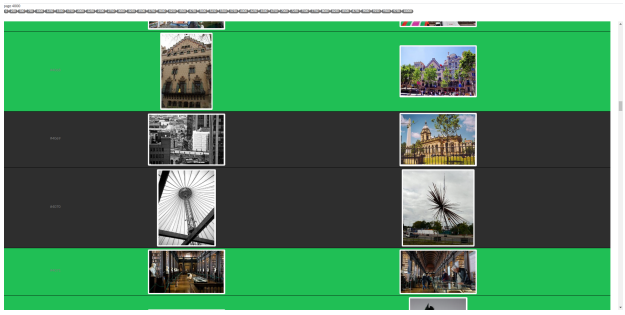

Figure 3: The annotation interface.

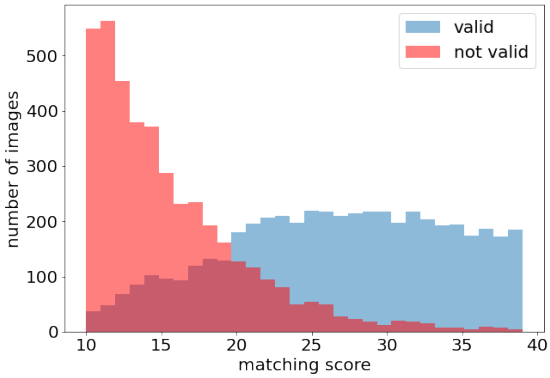

Figure 4: Distribution of the annotated image pairs

30 can be used to mitigate the number of false positives.

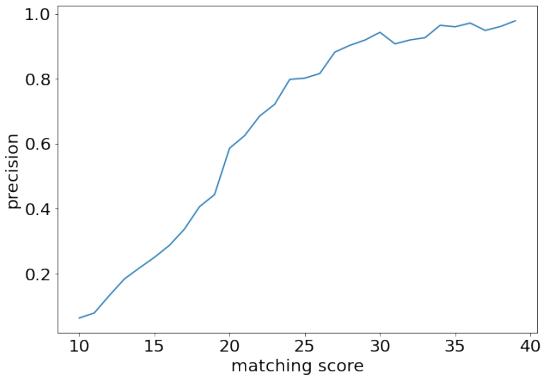

Figure 5: Distribution of the annotated image pairs

1.3. POI distribution

Figures 6 and 7 show the most present POIs and cities detected within Vis2Rec when all the post-processing and thresholding steps have been realized.

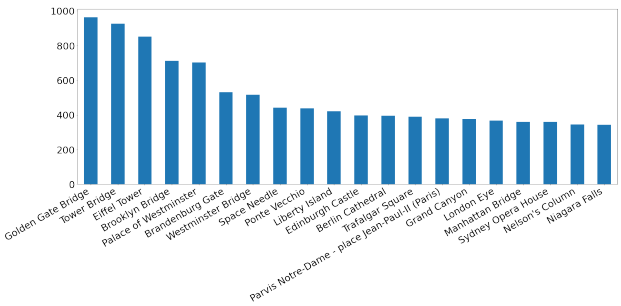

Figure 6: Most visited landmarks of Vis2Rec

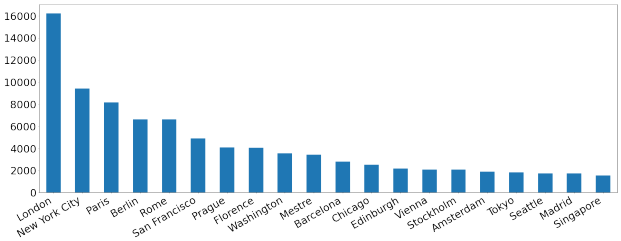

Figure 7: Most visited cities of Vis2Rec

2. POI recommentation

2.1. Methods

Hyperparameters of the used methods were optimized on the *validation* set. The results of this parameters search is presented in Table 1.

2.2. Results

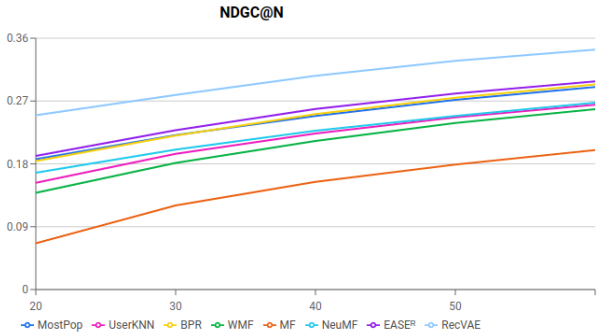

Figure 8: $NDGC@N$ of the test methods with respect to N

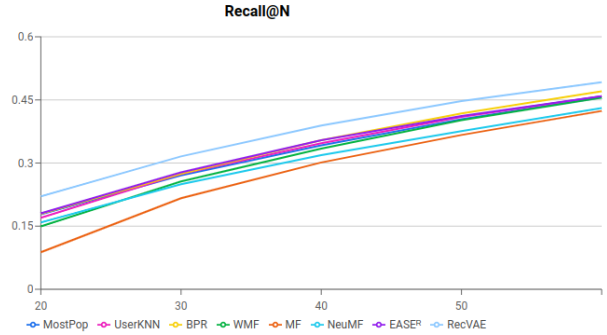

Figure 9: $Recall@N$ of the test methods with respect to N

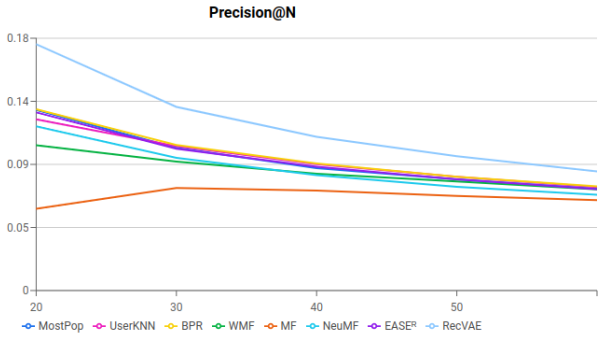

Figure 10: $Precision@N$ of the test methods with respect to N

Figures 8, 9, and 10 show respectively the $NDGC@N$, $recall@N$, and $precision@N$ with varying N. As we see, RecVAE clearly outperforms other methods for all metrics.

UserKNN	BPR	WMF	MF	NeuMF	EASE	RecVAE
$k = 500$	$k = 50$	$k = 100$	$k = 1000$	$batch\_size = 2048$	$\lambda = 500$	$hidden\_dim = 1000$
	$iters = 200$	$iters = 100$	$iters = 500$	$lr = 5 \times 10^{-4}$	$posB = True$	$latent\_dim = 1000$
	$lr = 10^{-2}$	$a = 1.0$	$lr = 5 \times 10^{-3}$	$n_{factors} = 8$		$\gamma = 0.01$
	$\lambda_{reg} = 0.01$	$b = 0.05$	$\lambda_{reg} = 0.01$	$layers = [64, 32, 16, 8]$		$batch\_size = 1024$
		$lr = 0.001$	$bias = False$	$n_{epochs} = 50$		$lr = 5 \times 10^{-4}$
		$\lambda_u = 0.01$		$n_{reg} = 10$		$n_{enc\_epochs} = 3$
		$\lambda_v = 0.01$				$n_{dec\_epochs=1}$
						$n_{epochs} = 30$
						$dropout = 0.5$

Table 1: Hyperparameters of the used methods