Supplementary Materials for
Noise-resistant Deep Metric Learning with Ranking-based Instance Selection

1. Generating Small Cluster Noise

To mimic characteristics of natural label noise, we propose a new model of noise synthesis called Small Cluster. In Algorithm 1, we show the pseudo-code for generating Small Cluster noise from a clean dataset. The algorithm first clusters images from a randomly selected ground-truth class into a large number of small clusters, using features extracted from a pretrained neural network. The number of clusters is set to $1/Z$ of the number of images in the class so each cluster is expected to have Z images. Each cluster is then merged into a randomly selected ground-truth class. The procedure is repeated until, out of the total of N images, the number of misplaced images reaches or exceeds the predefined percentage R.

In our experiments, we choose $Z = 2$ and set the random seed to 0. We use Mini-batch K-means [7] as our clustering algorithm.

Algorithm 1: Synthesizing Small Cluster Noise

Input : $\mathcal{X} = \{(x_0, \tilde{y_0}), (x_1, \tilde{y_1}), ..., (x_N, \tilde{y_N})\}$: training dataset
R: the noise rate
Z: mean number of images per cluster

Output : Y: the corrupted labels

```
Y = [ỹ_0, ỹ_1, ..., ỹ_N]
while #misplaced_images < RN do
    c = a uniformly sampled non-empty class
    X_c = {x_i | (x_i, ỹ_i) ∈ 𝒳, ỹ_i = c}
    Q = Clustering(X_c, n_cluster=int(|X_c|/Z))
    for q_i ∈ Q do
        c' = another uniformly sampled non-empty
             class that does not equal c
        for x_j ∈ q_i do
            Y[j] = c'
        end
    end
    Mark ∀x ∈ X_c as misplaced images
end
return Y
```

2. Training time on CARS and CUB.

Table 1 shows the training time with and without PRISM algorithm. PRISM adds about 100 seconds for 5K iterations, or 8% to 10% of total running time, on CARS and CUB.

Algorithm	CARS	CUB
MCL	1,170.05	1,189.91
MCL + PRISM	1,291.58 (+10.4%)	1,305.90 (+8.0%)
Soft Triple	1,186.658	1,184.75
Soft Triple + PRISM	1,279.97 (+7.9%)	1,284.97 (+10.2%)

Table 1: Training time (seconds) for 5K iterations.

3. Results on Landmark Recognition

We conduct experiments on landmark recognition datasets. We use Oxford [4] and Babenko's Landmark dataset [1] to train our model. RParis [6] is used to test the performance. Details of the dataset are described below.

- The **Oxford** Dataset [4] consists of 5,062 images of 11 Oxford landmarks, collected from Flickr. We utilize all the images (including images in which the buildings are not present, heavily occluded, or distorted).

- **Babenko's Landmark** Dataset [1] consists of 213,678 images of 672 landmarks. The images are retrieved by querying the Yandex image search engine with the name of landmarks. Certain level of label noise exists [2].

- The Revisited Paris (**RParis**) Dataset [6] contains 6,412 images of 12 landmarks in Paris. The dataset is originally created by [5] then cleaned by [6].

The training setting follows Section 4.3 of the main paper.

Results show that PRISM improves the performance on both small- and large-scale landmark recognition datasets with significant levels of label noise.

Table 2: mAP on RParis. Models are trained on Oxford Dataset.

Alorithm	Easy	Medium	Hard
MCL	60.8	47.9	24.8
MCL+PRISM	**61.7**	**48.8**	**25.7**
Soft Triple	63.9	49.9	25.2
Soft Triple+PRISM	**64.1**	**50.1**	**26.5**

Table 3: Precision@1 and MAP@R on RParis. Models are trained on Babenko's Landmark Dataset.

	P@1	MAP@R
MCL	82.04	21.80
MCL + PRISM (Ours)	**82.98**	**22.33**

4. Results on Clean Datasets

In Table 4, we report the results when the algorithm is trained on the original CUB, CARS, and SOP datasets. The training setting is identical to that in Section 4.3 of the main paper. The performance degradation on CUB is small. On SOP, filtering data at $R = 2\%$ and 5% causes performance to improve slightly. After inspection, we believe the original SOP dataset contains some noisy labels, indicating that noisy labels are common in real-world data.

Table 4: Precision@1 under different filtering rate R for MCL with PRISM.

Dataset	MCL only (R=0)	R=2%	R=5%	R=10%
CUB	60.8	60.4	60.0	60.1
CARS	82.1	81.3	80.2	79.3
SOP	81.0	81.2	81.1	80.8

5. Details of CARS-98N

Using the 98 labels from the CARS training set as the query terms, We build the CARS-98N dataset from the image search of Pinterest. No data cleaning has been performed. CARS-98N is only used for training. The size of CARS-98N is about 20% times larger than the training set of CARS [3] dataset.

Figure 1 shows the number of images for each class in CARS-98N dataset. It can be observed that the number of images is not evenly distributed across classes. Although many classes contain more than 100 images, fewer images can be found for certain car models such as Chevrolet Malibu Hybrid 2010, probably due to their limited market share or availability.

Figures 3 to 5 illustrate common types of noise in CARS-98N dataset. We take the class Dodge Durango SUV 2012 as an example. The noisy data contain images of different cars, as well as images of car parts and interior. As reference, we show the correct images in Figure 2. We also observe that for small classes, the noisy data are often unrelated to cars. For example, Figure 6 shows the noisy data in the class Eagle Talon Hatchback 1998.

References

[1] Artem Babenko, Anton Slesarev, Alexandr Chigorin, and Victor Lempitsky. Neural codes for image retrieval. In *ECCV*, pages 584–599. Springer, 2014. 1

[2] Albert Gordo, Jon Almazan, Jerome Revaud, and Diane Larlus. End-to-end learning of deep visual representations for image retrieval. *International Journal of Computer Vision*, 124(2):237–254, 2017. 1

[3] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *ICCV workshops*, pages 554–561, 2013. 2

[4] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *CVPR*, pages 1–8. IEEE, 2007. 1

[5] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *CVPR*, pages 1–8. IEEE, 2008. 1

[6] Filip Radenović, Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondřej Chum. Revisiting oxford and paris: Large-scale image retrieval benchmarking. In *CVPR*, pages 5706–5715, 2018. 1

[7] David Sculley. Web-scale k-means clustering. In *WWW*, pages 1177–1178, 2010. 1

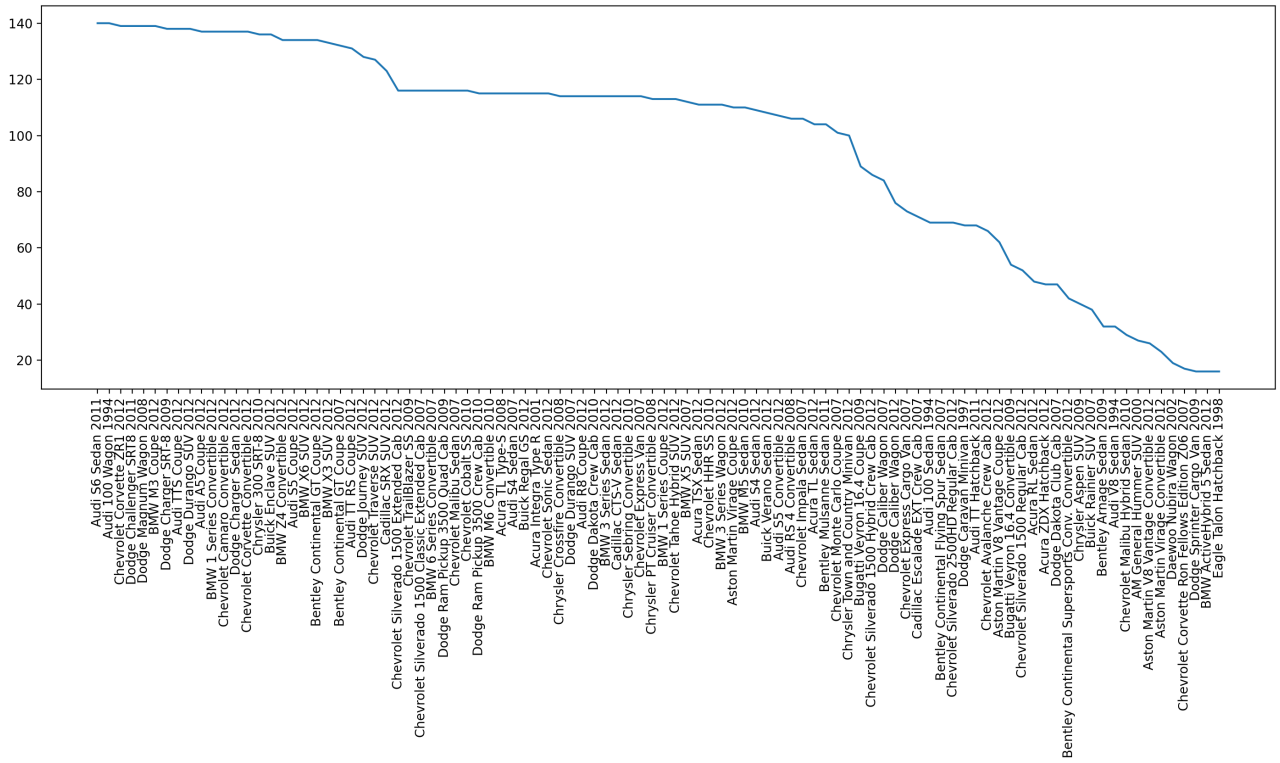

Figure 1: The number of images for each class in CARS-98N dataset. The X-axis gives the car model name and Y-axis refers to the number of images.

Figure 2: Images of Dodge Durango SUV 2012 that are correctly labeled in CARS-98N dataset.

Figure 3: Incorrect car models found in the class Dodge Durango SUV 2012 in CARS-98N.

Figure 4: Car part and accessory images in the class Dodge Durango SUV 2012 in the CARS-98N dataset.

Figure 5: Car interior images found in the class Dodge Durango SUV 2012 in CARS-98N.

Figure 6: Irrelevant images found in the class Eagle Talon Hatchback 1998 in CARS-98N.