Supplementary Material

1. Training details for baselines

For all the approaches we use ResNet-18 architectures and Adam optimizer with $\beta = [0.9, 0.99]$. For the cross-domain fine-tuning, we use the pre-training feature on dataset A as initialization.

1.1. One shot recognition baselines

For the Cosine Classifier entry and Matching Networks (scratch) entry, we trained our networks for 90 epochs with a learning rate of 1e-2 followed by 30 epochs with learning rate 1e-3. For the Matching Networks(fine-tune) entry and Weights Prediction entry we fine-tuned for 10 extra epochs using a learning rate of 1e-3. The Cosine Classifier is trained with batch size of 64 and the Matching Networks (scratch), Matching Networks(fine-tune), and Weights Prediction use a meta-batch size of 1 training episode which consists of 100 training examples and 100 test examples (for more details about the training procedure of meta-learning approaches we refer to [29] or their released code[1]).

1.2. Cross domain fine-tuning

We fix the learning rate as 1e-5 for all fine-tuning algorithms. For the triplet loss entry and our fine-tuning method with different τ, we follow the same training procedure : we use a batch-size of 8 and stop the training if the validation loss doesn't decrease for 10 epochs. For Nc-Net, we trained for 30 epochs with batch size 8.

2. Optimal parameters for feature comparison baselines

In Table 1, we provide a study on scales(number of feature in each dimension) and choices of features (Conv4 and Conv5) for all feature comparison baselines using synthetic references. To reproduce the following results, please refer to our released implementation[2]. We finally take image size 224 for all the baselines. The optimal features for different datasets are: *Conv5* feature for dataset A; *Conv4* feature for dataset B, Shoes and Chairs dataset.

Nb. Feature in Conv4 / Method	AvgPool		Concat		Local Simi.	
	Conv4	Conv5	Conv4	Conv5	Conv4	Conv5
10	5	13	57	18	57	18
12	4	11	62	22	64	21
14	2	**12**	**61**	29	**65**	23
16	3	9	59	27	65	22
18	2	8	59	26	64	20
20	2	8	57	29	61	21
22	2	5	56	27	60	21

Table 1: Influence of number of features in the reference image for all feature comparison baselines. We report top 1 accuracy in % on *dataset B Synthetic References* with using pre-training feature.

3. Dependency on τ for our fine-tuning algorithm

In this section, we provide a full analysis on the misalignment tolerance τ over different datasets: dataset B with engraving and synthetic references (Table 2); Shoes and Chairs datasets (Table 3)

[1] https://github.com/gidariss/FewShotWithoutForgetting
[2] https://github.com/XiSHEN0220/WatermarkReco/featComparisonBaseline

τ // Ref	Engraving	Synthetic
$\tau = 0$	67	73
$\tau = 1/22$	75	81
$\tau = 2/22$	73	79
$\tau = 3/22$	**75**	**83**
$\tau = 4/22$	73	77
$\tau = 5/22$	73	77
$\tau = \text{inf}$	65	77

Table 2: Effect of τ on the result of our spatial-aware fine-tuning strategy on *dataset B Synthetic and Engraving References*

τ // Ref	Chairs	Shoes
$\tau = 0$	89.7	40.9
$\tau = 1/24$	90.7	40.9
$\tau = 2/24$	89.7	44.3
$\tau = 3/24$	90.7	48.7
$\tau = 4/24$	**91.8**	**52.2**
$\tau = 5/24$	89.7	43.5
$\tau = \text{inf}$	89.7	47.0

Table 3: Effect of τ on the result of our spatial-aware fine-tuning strategy on *Shoes / Chairs dataset*

4. Evaluation on Briquet with engraving references

Method	Briquet-3b		Briquet-3b+Fine-tuning	
	acc.@1	acc.@1000	acc.@1	acc.@1000
AvgPool	0	9	0	19
Concat	19	72	22	78
Local Sim.	20	76	24	79
Ours N=1000	31	76	46	79

Table 4: Top-1 and top-1000 accuracy on our Briquet dataset with different models (”Briquet-3b” referred to using model trained on classification on dataset A and ”Briquet-3b+Fine-tuning” referred to using our fine-tuned model): the approaches are AvgPool, Concat, Local Similarity and first applying Local Similarity to obtain N = 1000 top ranked references then using our score to re-rank the N references.