

Appearance-Based Curriculum for Semi-Supervised Learning with Multi-Angle Unlabeled Data

Supplementary Material

Yuki Tanaka¹, Shuhei M. Yoshida¹, Takashi Shibata¹, Makoto Terao¹,
Takayuki Okatani^{2,3}, Masashi Sugiyama^{2,4}

¹NEC Corporation, ²RIKEN AIP, ³Tohoku University, ⁴The University of Tokyo

A. OpenCV implementation of d_{local}

Here, we present Python code to calculate d_{local} using OpenCV:

```
import cv2

def calc_similarity(image1, image2):
    akaze = cv2.AKAZE_create()
    matcher = cv2.BFMatcher(cv2.NORM_HAMMING)

    _, query_desc = akaze.detectAndCompute(
        image1, None
    )
    _, ref_desc = akaze.detectAndCompute(
        image2, None
    )

    matches = matcher.match(query_desc, ref_desc)

    dist_list = [it.distance for it in matches]
    dist = sum(dist_list) / len(dist_list)

    return dist
```

B. Details of experimental hyperparameters

To facilitate the reproduction of our results, we present the values of hyperparameters for each method and experimental setting.

All the settings have some hyperparameters in common: The choice of an optimization algorithm, learning rate, momentum, weight decay, training duration, batch size, and warm-up schedule and its duration. Due to the limitation on computational time, for each experimental setting, the optimizer, momentum, warm-up settings, and weight decay were fixed to predetermined values that were determined by preliminary experiments, as presented in Tab. 7. The batch size was chosen as the largest power of two that a GPU memory can hold, except for the experiments with DRINK, where it was set the smaller values accidentally. The learning rate and training duration were optimized using grid search. The selected values of these three hyperparameters are listed in Tab. 8.

There are also other algorithm-specific hyperparameters for semi-supervised learning (Tab. 9).

Hyperparameter	Value
Optimizer	SGD with Nesterov momentum
Momentum	0.9
Warm-up schedule	Linear
Warm-up duration	5% of the total iterations
Weight decay	5×10^{-4}
EMA	0.999

Table 7. Untuned hyperparameters common to all the experiments. We used the values listed here without tuning them for each setting. “Warm-up” refers to the learning-rate warm-up. “EMA” stands for the exponential moving average of model parameters, which was applied to all the semi-supervised settings.

For each experimental setting, we optimized the unlabeled loss coefficient for Π -model and VAT [21] and the threshold τ for FixMatch [27] and our method (ABC). The unlabeled loss coefficient weighs the loss function calculated from unlabeled data relative to the one calculated from labeled data. The threshold is used to select unlabeled data to which pseudo-labels are assigned. For Π -model, VAT, and FixMatch, the values of other hyperparameters were taken from the FixMatch paper [27] and the open source implementation¹.

Label propagation (LP) [36] has three hyperparameters: The scale parameter, threshold, and tolerance. The scale parameter σ is used to convert a pair-wise distance to the similarity. Specifically, given a distance $d(u, v)$ between two instances, u and v , their similarity $S(u, v)$ is calculated as

$$S(u, v) = \exp \left[-\frac{d(u, v)^2}{\sigma} \right]. \quad (6)$$

This is used as the weight of the edge between two nodes representing u and v . LP is essentially a diffusion process on the graph, and the tolerance is used to judge the convergence of LP. After this process, the unlabeled images with the prediction scores higher than the threshold are assigned pseudo-labels.

¹<https://github.com/TorchSSL/TorchSSL>

		θ_1	MIRO			θ_r	DRINK	COIL-100
			θ_2	θ_3				
Oracle	Iterations		30,000				8,000	10,000
	Learning rate		0.003				0.001	0.003
	Batch size		64				16	64
Supervised	Iterations	30,000	30,000	30,000	30,000	3,000	10,000	
	Learning rate	0.01	0.01	0.01	0.01	0.0001	0.03	
	Batch size	64	64	64	64	16	64	
Supervised (w/ rot. aug.)	Iterations	30,000	10,000	30,000	30,000	10,000	—	
	Learning rate	0.003	0.001	0.001	0.003	0.0001	—	
	Batch size	64	64	64	64	16	—	
LP (local desc.)	Iterations	10,000	30,000	30,000	30,000	10,000	—	
	Learning rate	0.01	0.01	0.01	0.003	0.0001	—	
	Batch size	64	64	64	64	32	—	
LP (global desc.)	Iterations	3,000	10,000	30,000	10,000	10,000	—	
	Learning rate	0.001	0.003	0.01	0.003	0.0001	—	
	Batch size	64	64	64	64	32	—	
II-model	Iterations	30,000	10,000	30,000	10,000	30,000	—	
	Learning rate	0.003	0.003	0.003	0.003	0.01	—	
	Batch size	16 / 16	16 / 16	16 / 16	16 / 16	16 / 16	—	
VAT	Iterations	10,000	10,000	30,000	10,000	30,000	—	
	Learning rate	0.003	0.01	0.01	0.003	0.01	—	
	Batch size	16 / 16	16 / 16	16 / 16	16 / 16	16 / 16	—	
FixMatch	Iterations	10,000	10,000	100,000	100,000	8,192	30,000	
	Learning rate	0.001	0.001	0.003	0.001	0.001	0.003	
	Batch size	8 / 24	8 / 24	8 / 24	8 / 24	4 / 28	8 / 24	
ABC (d_{angle})	Iterations	10,000	10,000	100,000	100,000	32,768	30,000	
	Learning rate	0.001	0.001	0.003	0.003	0.01	0.003	
	Batch size	8 / 24	8 / 24	8 / 24	8 / 24	16 / 16	16 / 16	
ABC (d_{local})	Iterations	10,000	10,000	100,000	10,000	48,000	30,000	
	Learning rate	0.001	0.001	0.003	0.003	0.01	0.003	
	Batch size	8 / 24	8 / 24	8 / 24	8 / 24	16 / 16	8 / 24	
ABC (d_{global})	Iterations	30,000	10,000	100,000	100,000	33,000	30,000	
	Learning rate	0.003	0.003	0.003	0.003	0.01	0.003	
	Batch size	8 / 24	8 / 24	8 / 24	8 / 24	16 / 16	8 / 24	

Table 8. Values of common hyperparameters: Training duration, learning rate, and batch size. We have only one oracle setting for MIRO because the oracle uses all the training data as labeled and conduct supervised learning. The notation “ B_L / B_U ” in the semi-supervised settings means that the batch sizes of labeled and unlabeled data are B_L and B_U , respectively.

		MIRO				DRINK	COIL-100
		θ_1	θ_2	θ_3	θ_r		
II-model	Unlabeled loss coeff.	1	10	10	10	10	—
VAT	Unlabeled loss coeff.	0.1	0.1	0.3	1	0.3	—
FixMatch	Threshold	0.9	0.9	0.9	0.9	0.9	0.9
ABC (d_{angle})	Threshold	0.9	0.7	0.9	0.7	0.9	0.5
ABC (d_{local})	Threshold	0.9	0.9	0.9	0.7	0.9	0.95
ABC (d_{global})	Threshold	0.7	0.7	0.9	0.9	0.9	0.9
LP (local desc.)	Scale	1,000	1,000	1,000	500	100	—
	Threshold	0.5	0.9	0.5	0.99	0.99	—
	Tolerance	0.0001	0.0001	0.0001	0.0001	0.0001	—
LP (global desc.)	Scale	0.0001	0.0001	3×10^{-5}	3×10^{-5}	1×10^{-5}	—
	Threshold	0.99	0.99	0.99	0.7	0.5	—
	Tolerance	0.0001	0.0001	0.0001	0.0001	0.0001	—

Table 9. Hyperparameters for the semi-supervised learning methods.