OSM: An Open Set Matting Framework with OOD Detection and Few-Shot Learning

Yuhongze Zhou[1]
yuhongze.zhou@mail.mcgill.ca
Issam Hadj Laradji[2]
issam.laradji@gmail.com
Liguang Zhou[3]
liguangzhou@link.cuhk.edu.cn
Derek Nowrouzezahrai[1]
derek@cim.mcgill.ca

[1] McGill University, Montreal, Canada
[2] ServiceNow Research
[3] The Chinese University of Hong Kong (Shenzhen), Shenzhen, China

Abstract

Natural image matting is the task of precisely estimating alpha mattes to separate foreground objects from background images. Existing matting methods only focus on classical closed-set problems where object categories and data distributions are similar between training and test sets. However, in the open world setup, there exists a situation where testing samples are drawn from a different distribution than the training data. To handle this situation, we present the first open set matting (OSM) framework that contains two networks: (1) an out-of-distribution (OOD) detection network to identify OOD to-be-matted objects; and (2) an incremental few-shot learning matting module to enlarge the existing knowledge base of to-be-matted objects. Our OOD detection network leverages metric-based prototype learning to be aware of unseen objects and increase inter-class separability, utilizing intra-batch connections to enhance intra-class compactness. Compared to other OOD detection methods, our network achieves state-of-the-art performance on SIMD dataset. Further, our incremental few-shot learning matting module improves the performance on unseen to-be-matted objects by gradually incorporating novel classes into the existing knowledge base without catastrophic forgetting and overfitting.

1 Introduction

The goal of natural image matting is to estimate alpha mattes to exactly extract foreground objects from background images. The matting problem can be formulated in a general mathematical manner that an image I is defined as a linear combination of alpha matte α, foreground F, and background B image,

$$I = \alpha F + (1-\alpha)B, \tag{1}$$

where RGB I is known, but F, B and α are unknown.

Apart from traditional matting approaches [1, 5, 7, 13, 14, 16, 24, 25, 26, 42, 45, 49, 50], deep learning has presented its powerful capability in matting tasks, which can be divided into three primary categories, including background-required [36, 40], only-image input [37, 52, 60], and trimap-needed [6, 8, 28, 31, 32, 46, 47, 54]. We focus on the most popular trimap-needed matting approaches where the trimap provides deterministic foreground, unknown, and background regions of an image. After Cho *et al*. [6] introduced deep neural networks into image matting task, Xu *et al*. [54] proposed a deep learning matting solution with a comprehensive matting database, known as the Adobe Image Matting dataset (AIM). Different from various matting works that emerged after Xu *et al*. work, recently, Sun *et al*. [46] has identified a bias issue of previous matting datasets, such as AIM [54] and the Distinctions-646 dataset [37]. To this end, they introduced a more balanced Semantic Image Matting Dataset (SIMD) as well as Semantic Image Matting network (SIM). The SIMD divides data into 20 different object categories according to object appearance within unknown region of trimap. Since the emergence of SIM, properly leveraging object information into matting task has caught researchers' interest. Although past matting methods have shown excellent performance in existing datasets, we notice that previous matting methods only focus on closed-set object categories whose performance can be degraded when encountering unseen objects. Therefore, we put matting into a real-world scenario and consider it as an open set task that is able to detect out-of-distribution (OOD) to-be-matted objects and find a matting performance balance between in-distribution (ID) and OOD to-be-matted objects.

Despite well-investigated open set learning [38], especially open set recognition (OSR), open set matting (OSM) remains an unexplored field. OSM is significantly valuable in practice because it makes detecting OOD to-be-matted objects possible, which can then be annotated by humans to obtain desirable results. However, challenges arise when dealing with OSM in the following aspects. First, in real applications, there can exist various kinds of matting objects that are unseen and challenging for the matting network trained on the closed-set knowledge base. Hence, identifying approaches for a closed-world discriminative model to be aware of unseen objects and training matting networks to mat new objects with a few labels are worth exploring and researching. Second, the network could suffer from interference of complex foreground and background information since it only detects whether the appearance of to-be-matted objects within unknown region of trimaps is OOD. Hence, without enhancing the expressiveness ability of the network, the capability of OOD detection could be degraded.

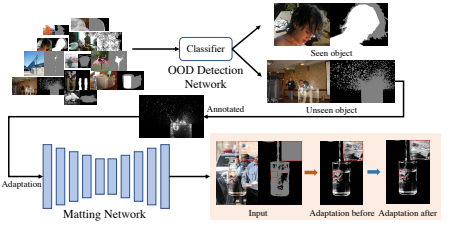

Figure 1: The overview of our open set matting (OSM) framework. The out-of-distribution (OOD) detection network detects unseen samples whose appearance within unknown region of trimap is unseen during training. After annotation of a few unseen samples, we conduct few-shot adaptation.

Therefore, in this paper, we propose the first open set matting framework: (1) an OOD detection network to identify OOD to-be-matted objects; (2) an incremental few-shot learning matting module to gradually enlarge the existing knowledge base of matting objects. To make the discriminative model trained on closed world to be unseen-aware and increase inter-class feature separability, we leverage metric-based prototype learning to embed samples into the low-dimensional prototype space. Further, to enhance the expressiveness ability of OOD detection network on matting data, we exploit metric-based intra-batch feature

connection to maintain intra-class feature compactness, in which "intra-batch" also means mini-batch while emphasizing that there is connectivity between the samples [39]. With these two carefully-designed components, our OOD detection network becomes unseen-aware and more adaptive to the matting task. Then, we adapt the matting network that is trained on closed-set data to unseen objects with only a few samples and without catastrophic forgetting/over-fitting. We compare our OOD detection network with other state-of-the-art OOD detection methods on SIMD dataset and show that our method obtains the new state-of-the-art results. We conduct experiments and analysis to validate the effectiveness of our few shot learning matting module.

To summarize, our contributions are as follows: **(1)** We propose the first open set matting (OSM) framework to tackle matting task from an open set perspective. **(2)** We show that our OOD detection network achieves the new state-of-the-art performance on SIMD dataset compared to other OOD detection methods. **(3)** We validate that our few-shot learning matting module can not only prevent catastrophic forgetting but also avoid over-fitting.

2 Related Work

2.1 Out-of-Distribution Detection

OOD detection can be categorized into two domains, i.e., uncertainty estimation-based and generative model-based approaches. For generative model-based OOD detection approaches, the network reconstruction error of ID data is smaller than that of OOD data. For uncertainty estimation-based OOD detection approaches, the maximum softmax probability (MSP) [18] is served as a baseline for uncertainty estimation. Hendrycks *et al*. [19] explore OOD detection in large-scale multi-class and multi-label settings and introduce maximum logit (MaxLogit) detector. However, one issue of MSP and MaxLogit is that DNNs tend to produce wrong prediction with high confidence in that DNNs are usually poorly calibrated [15]. Therefore, there are many works that aim to achieve better uncertainty estimation. For example, Guo *et al*. [15] evaluate various post-processing calibration methods and provide temperature scaling solution at calibrating predictions. Monte Carlo dropout (MC-dropout) [12] and ensembles [23] approaches leverage approximate Bayesian inference to better estimate uncertainty. Furthermore, Zaeemzadeh *et al*. [53] attempt to embed in-distribution data onto a union of 1-dimensional subspaces and leverage sampling-based approximate Bayesian inference for OOD detection (1D-subspaces).

2.2 Prototype Learning

Prototype learning is a deep learning counterpart of traditional nearest neighbor classification and Learning Vector Quantization (LVQ) [22] that relates each class to its corresponding prototype and conducts classification according to the distance based similarity between samples and prototypes. Prototype learning has demonstrated excellent performance in one-shot learning [11, 48], OOD/anomaly detection [3, 9, 55, 56], few-shot learning [35, 44, 51], and person re-identification [57]. It aims to learn a deep feature embedding whose semantic similarity possesses small intra-class variation but large inter-class variation. Since its goal also matches OOD detection task that there should be large inter-class gap between OOD data and ID data, we introduce prototype learning into our OOD detection network for open set matting.

3 Approach

3.1 Problem Setup

In Fig. 1, we provide the overview of our open set matting framework. This framework contains an OOD detection network and an incremental few-shot learning matting module. Consider that $I = \{I_1, I_2, ..., I_n\}$ are a set of images, $T = \{T_1, T_2, ..., T_n\}$ denote corresponding trimaps, and $A = \{\alpha_1, \alpha_2, ..., \alpha_n\}$ refer to corresponding alpha mattes. The closed-set data belongs to N ID classes $C_{\text{in}} = \{C_{\text{in},1}, ..., C_{\text{in},N}\}$ while K OOD classes, $C_{\text{out}} = \{C_{\text{out},1}, ..., C_{\text{out},K}\}$, are excluded from the closed-set data. Given I_i and T_i as input, the OOD detection network produces anomalous score S_{I_i} and identify OOD images by λ_{out} thresholding, that is, $I_i \in C_{\text{out}}$ (denoted as I_{out}) if $S_{I_i} > \lambda_{\text{out}}$, otherwise $I_i \in C_{\text{in}}$. Then, I_{out} would be forwarded to labelers who can provide the corresponding alpha matte A_{out}. With a few available samples of novel classes, the incremental few-shot learning matting module gradually enlarges the knowledge base of the closed-set matting network from C_{in} to $C_{\text{in}+K}$ where $C_{\text{in}+t} = C_{\text{in}} \cup \{C_{\text{out},1}, C_{\text{out},2}, ..., C_{\text{out},t}\}$, $t \in \{1, 2, ..., K\}$.

3.2 OOD Detection Network (OOD-DN)

Fig. 2 shows our OOD detection network that can be disentangled into a feature extractor and a discriminant function. The ResNet-50 [17] serves as a feature extractor $f(X;\theta_f)$ where X denotes the input image/trimap and θ_f serves as the parameters of feature extractor. The standard classification of DNNs is targeted to closed world that can be unsuitable for OOD detection. Hence, in order to increase the unseen-awareness and expressiveness of the network, we utilize prototype learning to build up distance features on top of the feature extractor. We calculate distances between the feature extractor output and predefined scaled one-hot prototypes to serve as the input of the discriminant function $g(\cdot)$ for classification [3, 53, 55]. Since prototypes are orthogonal to each other and prototypes can be easily extended to novel classes, it helps to increase inter-class separability and enable the network to be unseen-aware.

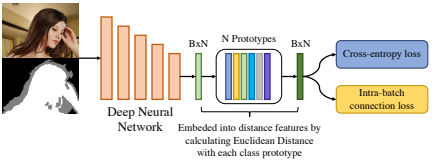

Figure 2: Our OOD Detection Network (OOD-DN). Our OOD-DN leverages prototype learning with intra-batch connection to be unseen-aware and generate informative logit features.

To be precise, consider all prototypes as $P = \{p_i \in \mathbb{R}^{1\times N} | i \in \{1, 2, ..., N\}\}$, where $p_i = [0, ..., \underset{i}{m}, ..., 0]$ corresponds to $C_{\text{in},i}$. We embed the latent feature output of the network that has the same length as the prototype into distance features by

$$d_i = -||f(X;\theta_f) - p_i||_2^2. \tag{2}$$

The final input feature for the discriminant function $g(\cdot)$ is formed by $D = \{d_i \in \mathbb{R} | i \in \{1, 2, ..., N\}\}$. Then, for classification, we optimize $f(X;\theta_f)$ and $g(\cdot)$ by minimizing the prototype learning based cross entropy loss, $\mathcal{L}_{\text{CE}}$. $\mathcal{L}_{\text{CE}}$ can be formulated as

$$\mathcal{L}_{\text{CE}} = -\log\left(\frac{\exp(d_y)}{\sum_{i=1}^{N}\exp(d_i)}\right), \tag{3}$$

where y is the ground-truth class label of input X and d_y refers to the distance feature between $f(X;\theta_f)$ and the prototype p_y. With prototype learning, we explicitly increase inter-class separability and enable the unseen-awareness of the network. Therefore, we expect our network to be more appropriate for OOD detection task.

Intra-Batch Connection Regularization In order to enhance the intra-class compactness and fully exploit data information we have, we leverage intra-batch connectivity, that is, for samples with the same label, their latent distance distributions should be similar while, for samples with different labels, their latent distance distributions should be distinguished. Therefore, we minimize Kullback-Leibler divergence between latent distance distributions of each pair of intra-batch samples that have the same class label over a total of N ID classes. The intra-batch connection loss $\mathcal{L}_{\text{IBC}}$ is defined as

$$\mathcal{L}_{\text{IBC}} = \sum_{i=1}^{N} \sum_{j=1, j<k}^{C_i} D_{\text{KL}}(D_{C_i}^{(j)} || D_{C_i}^{(k)}), \qquad D_{\text{KL}}(p||q) = \sum_{i=1}^{N} p_i \log\left(\frac{p_i}{q_i}\right), \tag{4}$$

where $D_{C_i}^{(m)} = \text{softmax}([d_1, d_2, ..., d_N])$ and C_i represents a cluster of samples that have the same label i.

OOD Detection During Inference Since the partition function constrains features to seen data and ignores unseen data, we use the negative maximum value of logit output D as the anomalous score for OOD detection without partition during inference [19]. Specifically, given input X, the anomalous score is defined as

$$S(f(X;\theta_f)) = -\max(d_i), \quad i \in \{1, 2, ..., N\}. \tag{5}$$

3.3 Incremental Few-Shot Learning Matting Module (IFL-MM)

The detected OOD to-be-matted images can then be delivered to labelers for annotation. We aim to enlarge the existing knowledge base of matting network to embrace novel classes without introducing external parameters and catastrophic forgetting under the situation where only a few labels are available. Thus, we introduce incremental few-shot learning matting module (IFL-MM) by (1) Train the matting network on 15-class ID data as the pre-trained model G; (2) Adapt the pre-trained model G to OOD domain with a few labels as the adapted model G'.

To illustrate our IFL-MM, we adopt U-Net architecture [28] as matting network. Considering the pre-trained matting network function $G(X;\theta)$ where θ is model parameters and X denotes input image and its corresponding trimap, we can obtain the estimated alpha matte $\hat{\alpha} = G(X;\theta)$. To improve OOD matting performance, first, we show that adapting the weights of matting network from ID domain to OOD domain directly by fine-tuning is inefficient since OOD data follows a different distribution than ID data and remodelling the statistics of Batch Normalization with exponential learning rate decay schedule can effectively handle this problem. Second, we quantify the importance of weights of matting network in ID domain for weight regularization of OOD data adaptation since a direct adaptation without regularization leads to over-fitting and knowledge forgetting.

Remodelling the Statistics of Batch Normalization with Exponential Learning Rate Decay In practice, we found that directly fine-tuning the pre-trained model on novel samples results in slow convergence and unstable training. It is due to the fact that (1) the traits of past data dominate over the statistics of Batch Normalization (BN) [20]; (2) the training can

be ill-conditioned if the feature transformation does not satisfy the condition of transforming inputs to be zero-mean, unit-variance, and uncorrelated [34, 53]; (3) when the existing knowledge base encounters novel samples, a non i.i.d. mini-batch situation arises and BN can fail.

Therefore, we propose to remodel the BN statistics with exponential learning rate decay to alleviate this issue. First, we are inspired from domain adaptation techniques, especially Adaptive Batch Normalization that recalculates the batch-wise mean and variance of BN at different layers of the network over the whole target domain before inference [27]. We reset the mean (resp. variance) of each BN of the pre-trained model to zero (resp. one) before fine-tuning. Upon resetting BN statistics and remodelling BN statistics as the running mean and variance over novel samples, we, to some extent, circumvent an non i.i.d. mini-batch situation and enable the network to obtain efficient adaptation ability. Second, to avoid unstable training, we exponentially decrease the learning rate $\hat{\eta}$ with respect to training iteration t, $\hat{\eta} = \eta_0 * \gamma^t$, where the initial learning rate η_0 is 0.01 and γ is the hyperparameter. The $\hat{\eta}$ will become 0.0001 after 3,000 iterations.

Weight Constrain by Synaptic Intelligence We argue that constraining the previous important model parameters can not only prevent over-fitting to limited samples but also avoid training collapse and divergence according to the following two reasons: (1) the direct fine-tuning without any regularization results in not only slow convergence but also over-fitting and catastrophic forgetting; (2) a matting network trained on ID data, different from well-investigated few-shot classification, can also be directly applied on OOD data although it turns out to be less accurate.

Elastic Weight Consolidation (EWC) [21] is a regularization technique that aims to overcome catastrophic forgetting by constraining the model parameters according to their importance for previous tasks. It uses Fisher information F to tell how much the model parameters θ_i commit to the observations. It can be achieved by adding an additional regularization term to the loss function when doing adaptation. Synaptic Intelligence (SI) [59] is a EWC simplified variant where F is calculated online by integrating the loss over the weight trajectories during gradient descent. We simplify the importance calculation by considering it as the expectation of the square of the partial derivative of the log-likelihood function with respect to θ_i. We minimize

$$L = |\alpha - \hat{\alpha}| + \frac{\lambda}{2} \cdot \sum_i F_i \cdot (\theta_i - \theta_i^*)^2, \qquad F_i = \mathbf{E}\left[\left(\frac{\partial}{\partial \theta_i} \mathcal{L}(\alpha|X;\theta)\right)^2\right], \tag{6}$$

where $\mathcal{L}(\alpha|X;\theta)$ is the log-likelihood function of previous tasks.

4 Experiments

4.1 OOD Detection Network (OOD-DN)

Datasets We conduct experiments on Semantic Image Matting Dataset (SIMD) that contains 20 classes with 726 training foregrounds and 89 testing foregrounds. To have a similar setup as Shaban *et al.* [41], we consider 5 classes, i.e., glass_ice, fire, water_drop, spider_web, and water_spray, out of 20 classes as OOD data and exclude these 5 classes from training set. During training, as commonly used with the SIMD dataset [46], we randomly composite training foregrounds with randomly selected background images from COCO [29]. For the

Methods	AUROC(IN)↑	AUPR(IN)↑	FPR95(IN)↓	AUROC(OUT)↑	AUPR(OUT)↑	FPR95(OUT)↓	DetectionError↓
MSP [18]	0.673	0.879	0.882	0.673	0.360	0.621	0.332
MaxLogit [19]	0.623	0.855	0.959	0.623	0.290	0.740	0.363
EnergyScore [30]	0.605	0.847	0.995	0.605	0.278	0.751	0.363
1-D Subspaces [58]	0.734	0.896	0.795	0.734	0.501	0.722	0.322
MMSP [3]	0.660	0.864	0.941	0.660	0.328	0.837	0.360
EDS [3]	0.630	0.810	0.959	0.630	0.319	1.000	0.367
OOD-DN (Ours)	**0.819**	**0.940**	**0.791**	**0.819**	**0.541**	**0.413**	**0.230**

Table 1: OOD detection results on SIMD dataset.

λ	PL	$\mathcal{L}_{CE}$	$\mathcal{L}_{IBC}$	MSP	MaxLogit	AUROC(IN)↑	AUPR(IN)↑	FPR95(IN)↓	AUROC(OUT)↑	AUPR(OUT)↑	FPR95(OUT)↓	DetectionError↓
	✓	✓			✓	0.315	0.663	0.996	0.315	0.183	0.964	0.485
	✓	✓		✓		0.664	0.857	0.841	0.664	0.353	0.722	0.349
$\lambda = 0.1$		✓	✓	✓		0.589	0.807	0.945	0.589	0.293	0.919	0.406
		✓	✓		✓	0.717	0.891	0.850	0.717	0.414	0.703	0.328
	✓	✓	✓	✓		0.493	0.738	0.955	0.493	0.247	0.979	0.464
	✓	✓	✓		✓	**0.819**	**0.940**	**0.791**	**0.819**	**0.541**	**0.413**	**0.230**
$\lambda = 1.0$		✓	✓	✓		0.763	0.917	0.923	0.763	0.431	0.576	0.283
		✓	✓		✓	0.752	0.917	0.950	0.752	0.390	0.558	0.279
	✓	✓	✓	✓		0.547	0.791	0.914	0.547	0.308	0.891	0.445
	✓	✓	✓		✓	0.743	0.876	0.655	0.743	0.573	0.848	0.287

Table 2: Ablation study results of our OOD detection network on SIMD dataset. PL refers to prototype learning.

test set, we follow Sun *et al.* [46] to synthesize 890 images that consist of 15 ID and 5 OOD classes. We also composite each SIMD training foreground with 10 randomly selected background images from COCO to synthesize 7,260 images as toy samples (denoted as Toy SIMD dataset). See additional results of another different OOD-ID split setting in the supplementary material.

Evaluation Metrics We evaluate OOD detection performance using the following metrics: (1) AUROC(IN): The area under the receiver operating characteristic; (2) AUPR(IN): The area under the precision-recall curve; (3) FPR95(IN): The false positive rate at 95% true positive rate; (4) AUROC(OUT); (5) AUPR(OUT); (6) FPR95(OUT); (7) Detection Error that indicates the minimum misclassification probability. Metrics suffixed by **(IN)** are calculated when ID data is treated as positive. Opposite to **(IN)**, metrics suffixed by **(OUT)** are calculated when OOD data is treated as positive.

Implementation Details We follow similar data processing and augmentation procedure as GCA-Matting [28] to generate random trimaps and augmented images. We randomly crop square patches from the unknown region of composited images and then resize them to 320×320 patches. The network is trained for 50,000 iterations with 20 batch size. The Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ is adopted with initialized learning rate, 0.01, plus warm-up and cosine decay techniques. The hyperparameter of $\mathcal{L}_{IBC}$, λ is set to 0.1 and m is 3.

Results We compare our OOD-DN results with MSP [18], MaxLogit [19], energy score [30], 1D-subspaces[1] [58], metric-based maximum softmax probability (MMSP) [3], and Euclidean distance sum (EDS) [3]. We reproduce these methods according to their official public available code under the same training configuration as ours. In Table 1, we present the quantitive comparison. The results show that our OOD-DN achieves the state-of-the-art performance in all OOD detection related metrics. Compared to the state-of-the-art 1D-subspaces, our OOD-DN outperforms it with a relevantly big margin especially in AUROC(IN), AUROC(OUT),

[1]For 1D-subspaces, the first singular vector of each class is calculated by using the extracted features from Toy SIMD dataset of the corresponding class.

Methods	SAD(IN)↓	MSE(IN)↓	Grad(IN)↓	Conn(IN)↓	SAD(OUT)↓	MSE(OUT)↓	Grad(OUT)↓	Conn(OUT)↓
Pre-trained	33.71	9.7	18.62	29.99	79.47	16.2	51.44	77.47
Finetune	154.07±17.45	146.2±23.0	124.52±9.68	161.36±17.52	147.46±9.86	72.9±7.2	144.12±20.07	149.82±10.54
IFL-MM (Ours)	**44.87±5.13**	**17.8±4.8**	**24.17±2.40**	**43.74±5.86**	**68.08±3.56**	**14.8±1.1**	**44.41±3.34**	**63.35±4.22**
OSM (Ours)	**37.22±2.54**	**13.41±2.73**	**20.60±1.76**	**35.06±3.06**	**70.78±3.84**	**14.73±1.05**	**46.64±2.80**	**66.93±4.40**

Table 3: Matting results on SIMD dataset.

Classes	defocus	fur	hair_easy	hair_hard	insect	motion	net	flower	leaf	tree
Pre-trained	12.75	8.02	6.98	10.90	120.34	4.73	75.15	49.75	34.27	70.25
Finetune	183.11	36.02	41.19	53.51	278.94	25.45	267.20	204.39	187.52	310.73
IFL-MM (Ours)	**55.74**	**9.53**	**8.74**	**12.54**	**103.27**	**6.15**	**103.04**	**64.84**	**37.70**	**95.01**
OSM (Ours)	**44.95**	**8.62**	**7.32**	**11.44**	**90.42**	**5.37**	**85.12**	**57.63**	**33.99**	**76.33**
Classes	plastic_bag	sharp	smoke_cloud	lace	silk	**glass_ice**	**fire**	**water_drop**	**spider_web**	**water_spray**
Pre-trained	32.09	2.43	28.78	75.07	60.65	92.00	80.63	40.60	162.78	46.88
Finetune	295.79	26.54	263.50	470.61	231.77	219.07	127.99	70.19	234.73	104.76
IFL-MM (Ours)	**85.82**	**2.67**	**53.39**	**106.31**	**84.32**	**89.35**	**69.18**	**30.20**	**128.80**	**39.58**
OSM (Ours)	**42.47**	**2.61**	**33.42**	**92.41**	**77.96**	**89.08**	**79.21**	**31.31**	**133.16**	**38.18**

Table 4: Detailed quantitive matting results of 20 classes of SIMD dataset on SAD metric. Bolden classes are OOD classes, otherwise classes are ID classes.

Reg	ExpDecay	RemodelBN	SAD(IN)↓	MSE(IN)↓	Grad(IN)↓	Conn(IN)↓	SAD(OUT)↓	MSE(OUT)↓	Grad(OUT)↓	Conn(OUT)↓
✓	✓		**43.89**	**16.0**	24.27	**42.54**	69.64	**14.8**	**44.37**	65.34
✓		✓	47.65	18.9	25.23	46.56	69.25	15.5	46.66	64.98
	✓	✓	153.54	142.4	118.98	160.43	149.15	74.5	141.17	151.24
✓	✓	✓	44.87	17.8	**24.17**	43.74	**68.08**	**14.8**	44.41	**63.35**

Table 5: Ablation study results of our incremental few-shot learning matting module on SIMD dataset. Note that **Reg** is the regularization term based on Elastic Weight Consolidation (EWC).

FPR95(OUT), and Detection Error metrics. Besides, our OOD-DN does not require a time-consuming sampling procedure like 1D-subspaces.

Ablation Study We present ablation experimental results of our OOD-DN to study the effect of different hyperparameters of $\mathcal{L}_{\text{IBC}}$, loss functions, and OOD inference strategies, as shown in Table 2. We can see that when prototype learning (PL) incorporates with intra-batch connection regularization, the network can produce informative logit features for OOD detection, thus showing the effectiveness of MaxLogit. Furthermore, the ablation study shows that, by utilizing either PL or $\mathcal{L}_{\text{IBC}}$, the network can improve most of OOD detection metrics compared to baselines, which demonstrates the superiority of our OOD-DN.

4.2 Incremental Few-Shot Learning Matting Module (IFL-MM)

Datasets In the initial training stage, we utilize 15-class ID data of SIMD training set as training data. In the adaptation stage, we randomly sample 5-way 6-shot images from 5-class OOD data of Toy SIMD dataset as training set and consider the remaining images of 5-class OOD data of Toy SIMD dataset as validation set by excluding images whose foregrounds are overlapped with that of training set. We leverage SIMD test set as test set.

Evaluation Metrics We use common matting evaluation metrics, i.e., Sum of Absolute Differences (SAD), Mean Squared Error (MSE) that is $\times 10^3$, Gradient error (Grad), and Connectivity error (Conn), to evaluate matting performance. Metrics suffixed by **(IN)** (resp. **(OUT)**) are calculated on ID (resp. OOD) data. We conduct few-shot adaptation exper-

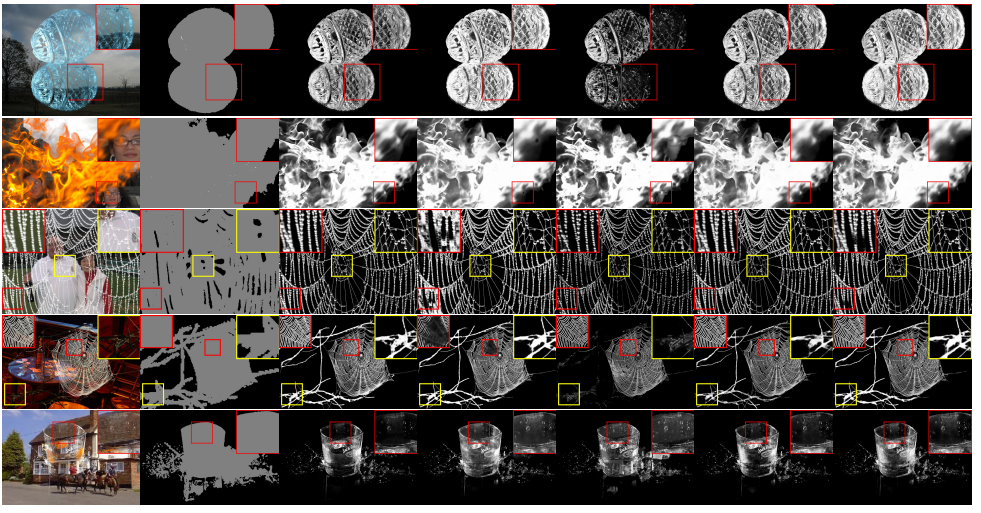

Figure 3: Visual comparison of matting results on 5 OOD classes of SIMD dataset. From the 1st row to the 5th row, glass_ice, fire, water_drop, spider_web, and water_spray. From left to right, image, trimap, GT, Pre-trained model, Finetune, IFL-MM (Ours), and OSM (Ours).

iments 10 times and report average metrics and their corresponding standard deviations. **Implementation Details** In the initial training stage, we adopt the similar training strategy as GCA-Matting [28] and use Adam optimizer to train our matting network on ID data with 20 batch size, 200,000 iterations, and $4e$-4 initialized learning rate. In the adaptation stage, we perform various data augmentation techniques before composition and random 512×512 patch cropping. Specifically, for each image, we apply random scaling, horizontal flipping, rotation, and color jittering. For trimap generation, we erode and dilate alpha matte with a random kernel size within $[1, 29]$ respectively. To extend limited data, we randomly merge the foreground of another randomly selected image with the current image foreground. The network is trained for 3,000 iterations with 20 batch size. The Adam optimizer with $\beta_1 = 0.9999$ and $\beta_2 = 0.9999$ is used with initialized learning rate, 0.01. The exponential decay schedule of learning rate is utilized. The λ is set to $2e8$.

Results We compare our IFL-MM with pre-trained and fine-tuned models. The Table 3 presents quantitive results of ID and OOD domains. The results show that our method improves performance on OOD data by a big margin, especially in SAD, Grad, and Conn metrics. Besides, unlike the fine-tuned model that nearly forgets ID data, our method successfully alleviates catastrophic forgetting about ID data. The tremendous performance gap between the fine-tuned model and ours demonstrates that the direct fine-tuning results in slow convergence and inefficiency in both time and ID/OOD data performance. Further, we present quantitive results of 20 classes on SAD metrics in Table 4. Our method outperforms the pre-trained model in every OOD class and surpasses the fine-tuned model in all 20 classes. Besides, we found that our IFL-MM is sensitive to datasets since the pre-trained model can generalize well with unseen categories whose correlations are close to training data. To better illustrate the superiority of our IFL-MM, we compare the visual matting results between baselines and ours in Fig. 3 on 5 challenging OOD classes. Our IFL-MM can better separate target objects from background without apparent background ghost or miss-

ing foreground details and obtain overall visual improvement compared to both pre-trained and fine-tuned models.

Ablation Study We conduct ablation study on our IFL-MM to investigate the effectiveness of each component. In Table 5, we compares our IFL-MM with IFL-MM without regularization (Reg), remodelling BN statistics (RemodelBN), and exponential learning rate decay (ExpDecay). The results show that each component has its own contribution to OOD data adaptation. To further demonstrate the effectiveness of RemodelBN and ExpDecay, in Fig. 5, we present the curve comparison of validation performance during training. The IFL-MM convergence speed is the best compared to IFL-MM without RemodelBN/ExpDecay and, in the end, the IFL-MM validation performance is on par with or even better than the other two. Further, as indicated in Table 5, our test performance in OOD (resp. ID) data is overall better than (resp. comparable with) IFL-MM without Reg/RemodelBN/ExpDecay. Therefore, the above observations demonstrate the faster convergence speed and anti-over-fitting ability of our IFL-MM.

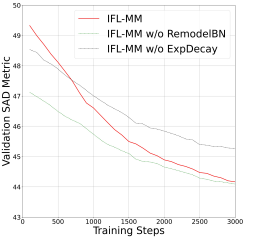

Figure 4: Comparison of training process among IFL-MM w/o RemodelBN, IFL-MM w/o ExpDecay, and IFL-MM on SAD metric of validation set.

4.3 Open Set Matting (OSM)

We build our open set matting (OSM) framework by the following steps: (1) Train our OOD-DN on 15 ID class data; (2) Obtain detected OOD data out of the Toy SIMD dataset; (3) Adapt the pre-trained matting network to OOD data by leveraging our IFL-MM. Noted that, instead of 5-way 6-shot images, we randomly sample 30 images out of detected OOD data. Our OSM results are shown in Table 3. The detailed results of each class are presented in Table 4. It is obvious that, in challenging 5 OOD classes, our OSM is competitive against IFL-MM trained with purely OOD samples. We present the visual results of our OSM in Fig. 3. It is noted that our OSM significantly improves OOD matting visual results and sometimes is on par with IFL-MM that is trained by purely OOD data.

To make our open set matting framework progressively incorporate novel classes, our OOD-DN can be combined with research about open world recognition [2, 9, 10, 43] to scale elegantly with the increasing number of classes. Then, the cycle of our open set matting framework can be pushed to open world matting.

5 Conclusion and Future Work

We introduce the first open set matting (OSM) framework that contains two networks, an OOD detection network (OOD-DN) and an incremental few-shot learning matting module (IFL-MM). Our OOD-DN leverages prototype learning and intra-batch connection to be aware of unseen objects, maintain inter-class separability and intra-class compactness, and achieve the state-of-the-art OOD detection performance. Our IFL-MM can effectively prevent catastrophic forgetting and over-fitting. For future work, our OOD-DN and IFL-MM still have performance improvement space and our OSM can be extended to scale flexibly with the increasing number of classes for open world matting.

Acknowledgements We thank ServiceNow for providing GPU resources for this project with the ServiceNow Toolkit. Thanks to Pierre-André Noël for feedback on the paper draft.

References

[1] Yagiz Aksoy, Tunc Ozan Aydin, and Marc Pollefeys. Designing effective inter-pixel information flow for natural image matting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 29–37, 2017.

[2] Abhijit Bendale and Terrance Boult. Towards open world recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1893–1902, 2015.

[3] Jun Cen, Peng Yun, Junhao Cai, Michael Yu Wang, and Ming Liu. Deep metric learning for open world semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15333–15342, 2021.

[4] Guangyao Chen, Limeng Qiao, Yemin Shi, Peixi Peng, Jia Li, Tiejun Huang, Shiliang Pu, and Yonghong Tian. Learning open set network with discriminative reciprocal points. In *European Conference on Computer Vision*, pages 507–522. Springer, 2020.

[5] Qifeng Chen, Dingzeyu Li, and Chi-Keung Tang. Knn matting. *IEEE transactions on pattern analysis and machine intelligence*, 35(9):2175–2188, 2013.

[6] Donghyeon Cho, Yu-Wing Tai, and Inso Kweon. Natural image matting using deep convolutional neural networks. In *European Conference on Computer Vision*, pages 626–643. Springer, 2016.

[7] Yung-Yu Chuang, Brian Curless, David H Salesin, and Richard Szeliski. A bayesian approach to digital matting. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 2, pages II–II. IEEE, 2001.

[8] Yutong Dai, Hao Lu, and Chunhua Shen. Learning affinity-aware upsampling for deep image matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6841–6850, 2021.

[9] Rocco De Rosa, Thomas Mensink, and Barbara Caputo. Online open world recognition. *arXiv preprint arXiv:1604.02275*, 2016.

[10] Tri Doan and Jugal Kalita. Overcoming the challenge for text classification in the open world. In *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 1–7. IEEE, 2017.

[11] Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, 28(4):594–611, 2006.

[12] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[13] Eduardo SL Gastal and Manuel M Oliveira. Shared sampling for real-time alpha matting. In *Computer Graphics Forum*, volume 29, pages 575–584. Wiley Online Library, 2010.

[14] Leo Grady, Thomas Schiwietz, Shmuel Aharon, and Rüdiger Westermann. Random walks for interactive alpha-matting. In *Proceedings of VIIP*, volume 2005, pages 423–429, 2005.

[15] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International Conference on Machine Learning*, pages 1321–1330. PMLR, 2017.

[16] Kaiming He, Christoph Rhemann, Carsten Rother, Xiaoou Tang, and Jian Sun. A global sampling method for alpha matting. In *CVPR 2011*, pages 2049–2056. IEEE, 2011.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[18] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *Proceedings of International Conference on Learning Representations*, 2017.

[19] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joe Kwon, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. Scaling out-of-distribution detection for real-world settings. *ICML*, 2022.

[20] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[21] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

[22] Teuvo Kohonen. Learning vector quantization. In *Self-organizing maps*, pages 175–189. Springer, 1995.

[23] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

[24] Philip Lee and Ying Wu. Nonlocal matting. In *CVPR 2011*, pages 2193–2200. IEEE, 2011.

[25] Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *IEEE transactions on pattern analysis and machine intelligence*, 30(2):228–242, 2007.

[26] Anat Levin, Alex Rav-Acha, and Dani Lischinski. Spectral matting. *IEEE transactions on pattern analysis and machine intelligence*, 30(10):1699–1712, 2008.

[27] Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. Revisiting batch normalization for practical domain adaptation. *arXiv preprint arXiv:1603.04779*, 2016.

[28] Yaoyi Li and Hongtao Lu. Natural image matting via guided contextual attention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11450–11457, 2020.

[29] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[30] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems*, 33:21464–21475, 2020.

[31] Hao Lu, Yutong Dai, Chunhua Shen, and Songcen Xu. Indices matter: Learning to index for deep image matting. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3266–3275, 2019.

[32] Sebastian Lutz, Konstantinos Amplianitis, and Aljosa Smolic. Alphagan: Generative adversarial networks for natural image matting. *arXiv preprint arXiv:1807.10088*, 2018.

[33] Dimity Miller, Niko Sunderhauf, Michael Milford, and Feras Dayoub. Class anchor clustering: A loss for distance-based open set recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3570–3578, 2021.

[34] Grégoire Montavon, Geneviève Orr, and Klaus-Robert Müller. *Neural networks: tricks of the trade*, volume 7700. springer, 2012.

[35] Boris Oreshkin, Pau Rodríguez López, and Alexandre Lacoste. Tadam: Task dependent adaptive metric for improved few-shot learning. *Advances in neural information processing systems*, 31, 2018.

[36] R. J. Qian and M. I. Sezan. Video background replacement without a blue screen. In *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, volume 4, pages 143–146 vol.4, 1999.

[37] Yu Qiao, Yuhao Liu, Xin Yang, Dongsheng Zhou, Mingliang Xu, Qiang Zhang, and Xiaopeng Wei. Attention-guided hierarchical structure aggregation for image matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13676–13685, 2020.

[38] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1757–1772, 2012.

[39] Jenny Denise Seidenschwarz, Ismail Elezi, and Laura Leal-Taixé. Learning intra-batch connections for deep metric learning. In *International Conference on Machine Learning*, pages 9410–9421. PMLR, 2021.

[40] Soumyadip Sengupta, Vivek Jayaram, Brian Curless, Steven M Seitz, and Ira Kemelmacher-Shlizerman. Background matting: The world is your green screen. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2291–2300, 2020.

[41] Amirreza Shaban, Shray Bansal, Zhen Liu, Irfan Essa, and Byron Boots. One-shot learning for semantic segmentation. In *BMVC*, 2017.

[42] Ehsan Shahrian, Deepu Rajan, Brian Price, and Scott Cohen. Improving image matting using comprehensive sampling sets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 636–643, 2013.

[43] Yu Shu, Yemin Shi, Yaowei Wang, Yixiong Zou, Qingsheng Yuan, and Yonghong Tian. Odn: Opening the deep network for open-set action recognition. In *2018 IEEE international conference on multimedia and expo (ICME)*, pages 1–6. IEEE, 2018.

[44] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

[45] Jian Sun, Jiaya Jia, Chi-Keung Tang, and Heung-Yeung Shum. Poisson matting. In *ACM SIGGRAPH 2004 Papers*, pages 315–321. 2004.

[46] Yanan Sun, Chi-Keung Tang, and Yu-Wing Tai. Semantic image matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11120–11129, 2021.

[47] Jingwei Tang, Yagiz Aksoy, Cengiz Oztireli, Markus Gross, and Tunc Ozan Aydin. Learning-based sampling for natural image matting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3055–3063, 2019.

[48] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29, 2016.

[49] Jue Wang and Michael F Cohen. An iterative optimization approach for unified image segmentation and matting. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 936–943. IEEE, 2005.

[50] Jue Wang and Michael F Cohen. Optimized color sampling for robust matting. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[51] Yu-Xiong Wang, Ross Girshick, Martial Hebert, and Bharath Hariharan. Low-shot learning from imaginary data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7278–7286, 2018.

[52] Tianyi Wei, Dongdong Chen, Wenbo Zhou, Jing Liao, Hanqing Zhao, Weiming Zhang, and Nenghai Yu. Improved image matting via real-time user clicks and uncertainty estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15374–15383, 2021.

[53] Simon Wiesler and Hermann Ney. A convergence analysis of log-linear training. *Advances in Neural Information Processing Systems*, 24, 2011.

[54] Ning Xu, Brian Price, Scott Cohen, and Thomas Huang. Deep image matting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2970–2979, 2017.

[55] Hong-Ming Yang, Xu-Yao Zhang, Fei Yin, and Cheng-Lin Liu. Robust classification with convolutional prototype learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3474–3482, 2018.

[56] Hong-Ming Yang, Xu-Yao Zhang, Fei Yin, Qing Yang, and Cheng-Lin Liu. Convolutional prototype network for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[57] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Deep metric learning for person re-identification. In *2014 22nd international conference on pattern recognition*, pages 34–39. IEEE, 2014.

[58] Alireza Zaeemzadeh, Niccolò Bisagno, Zeno Sambugaro, Nicola Conci, Nazanin Rahnavard, and Mubarak Shah. Out-of-distribution detection using union of 1-dimensional subspaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9452–9461, 2021.

[59] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *International Conference on Machine Learning*, pages 3987–3995. PMLR, 2017.

[60] Yunke Zhang, Lixue Gong, Lubin Fan, Peiran Ren, Qixing Huang, Hujun Bao, and Weiwei Xu. A late fusion cnn for digital matting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7469–7478, 2019.