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

Yuhongze Zhou, School of Computer Science

McGill University, Montreal

Nov., 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Computer Science

under the supervision of Dr. Issam Hadj Laradji,

Prof. Derek Nowrouzezahrai.

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.

Abrégé

Le matage naturel des images consiste à estimer avec précision les caches alpha pour séparer les objets de premier plan des images d'arrière-plan. Les méthodes de matage existantes se concentrent uniquement sur les problèmes classiques en ensembles fermé où les catégories d'objets et les distributions des données sont similaires entre les ensembles d'apprentissage et de test. Cependant, dans la configuration du monde ouvert, il existe une situation où les échantillons de test sont tirés d'une distribution différente de celle des données d'apprentissage. Pour gérer cette situation, nous présentons le premier cadre de matage en ensemble ouvert (OSM) qui contient deux réseaux : (1) un réseau de détection hors distribution (OOD) pour identifier les objets OOD à mater ; et (2) un module de matage d'apprentissage incrémental en apprentissage *few shot* pour élargir la base de connaissances existante des objets à mater. Notre réseau de détection OOD s'appuie sur l'apprentissage de prototypes basé sur la métrique pour être conscient des objets inconnus et augmenter la séparabilité inter-classes en utilisant des connexions intra-lots pour améliorer la compacité intra-classe. Comparé à d'autres méthodes de détection OOD, notre réseau atteint des performances à l'état de l'art sur l'ensemble de données SIMD. De plus, notre module de matage d'apprentissage incrémentiel en apprentissage *few shot* améliore les performances sur les objets invisibles matage en incorporant progressivement de nouvelles classes dans la base de connaissances existante sans oubli catastrophique ni surapprentissage.

Acknowledgements

I wish to thank everyone who made this project possible. I would like to thank Liguang Zhou for inspiring me to conceive this initial project direction and express my gratitude to my supervisor, Prof. Derek Nowrouzezahrai, for allowing me to give this project a shot with support. I also thank Dr. Issam Hadj Laradji for his patient and detailed guidance throughout the course of this project. Besides, thank ServiceNow for providing GPU resources for this project with the ServiceNow Toolkit. Thanks to Pierre-André Noël for feedback on the manuscript draft. Thanks to Pierre-André Noël, Fanny Rancourt, and Marilyse Solis for comments on French abstract.

Contribution of Authors

I completed this thesis on my own. This thesis is based on a joint work with Issam Hadj Laradji, Liguang Zhou, and Derek Nowrouzezahrai, appearing at the BMVC 2022 conference [203] as a spotlight. The paper link is `https://bmvc2022.mpi-inf.mpg.de/0092.pdf`. Inspired by Liguang Zhou, I proposed this topic. Prof. Derek Nowrouzezahrai introduced Dr. Issam Hadj Laradji to help me. I came up with the idea, conducted experiments, and wrote the manuscript. Dr. Issam Hadj Laradji guided me throughout this project, taught me how to write, polished the manuscript a bit, and taught/helped me do the rebuttal.

Table of Contents

Abstract i
Abrégé ii
Acknowledgements iii
Contribution of Authors iv
List of Figures viii
List of Tables ix

1 Introduction 1

2 Related Work 5

2.1 Natural Image Matting 5
2.2 Out-of-Distribution (OOD) Detection 8
2.3 Prototype Learning 9
2.4 Incremental Learning 10
2.5 Open Set Domain Adaptation 12
2.6 Open Set Learning 14
2.6.1 Basic Notations and Definitions 14
2.6.2 Open Set Recognition (OSR) 16

3 Approach 20

3.1 Problem Setup 20
3.2 OOD Detection Network (OOD-DN) 21
3.2.1 Intra-Batch Connection Regularization 22

3.2.2 OOD Detection During Inference 23
3.3 Incremental Few-Shot Learning Matting Module (IFL-MM) 23
3.3.1 Remodelling the Statistics of Batch Normalization with Exponential Learning Rate Decay 24
3.3.2 Weight Constrain by Synaptic Intelligence 25

4 Experiments **27**
4.1 OOD Detection Network (OOD-DN) 27
4.1.1 Datasets 27
4.1.2 Evaluation Metrics 28
4.1.3 Implementation Details 28
4.1.4 Results 29
4.1.5 Ablation Study 29
4.2 Incremental Few-Shot Learning Matting Module (IFL-MM) 30
4.2.1 Datasets 30
4.2.2 Evaluation Metrics 30
4.2.3 Implementation Details 30
4.2.4 Results 31
4.2.5 Ablation Study 32
4.3 Open Set Matting (OSM) 33

5 Discussion **35**
5.1 Out-of-distribution Class Selection 35
5.2 Limitations 36
5.3 Future Works 37
5.3.1 Dataset 37
5.3.2 Single-type Object vs. Multiply-type Objects 38
5.3.3 Trimap-needed Matting vs. Trimap-free Matting 38
5.3.4 Open World Matting 39

6 Conclusions **40**

7 Appendix **41**

List of Figures

1.1 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. 2

3.1 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. 21

3.2 The pipeline of our incremental few-shot learning matting module (IFL-MM). After training the matting network on ID data, we extend it to OOD data with a few samples but without catastrophic forgetting. 23

4.1 Comparison of training process among IFL-MM w/o RemodelBN, IFL-MM w/o ExpDecay, and IFL-MM on SAD metric of validation set. 33

4.2 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). 34

List of Tables

4.1 OOD detection results on SIMD dataset. 28
4.2 Ablation study results of our OOD detection network on SIMD dataset. PL refers to prototype learning. 28
4.3 Matting results on SIMD dataset. 31
4.4 Detailed quantitive matting results of 20 classes of SIMD dataset on SAD metric. Bolden classes are OOD classes, otherwise classes are ID classes. . . 31
4.5 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). 32

7.1 Additional OOD detection results on SIMD dataset. 41

Chapter 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,

$$I = \alpha F + (1 - \alpha)B, \tag{1.1}$$

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

Apart from traditional matting approaches [4,28,31,52,57,61,80–82,146,153,166,167], deep learning has presented its powerful capability in matting tasks, which can be divided into three primary categories, including background-required [89, 124, 143, 179], only-image input [75,83,85,93,125,170,200], and trimap-needed [29,35,86,94,102,104,121, 156, 157, 177]. 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.* [29] introduced deep neural networks into image matting task, Xu *et al.* [177] proposed a deep learning matting solution with a comprehensive matting database, also known as the Adobe Image Matting dataset (AIM). Different from various matting works that emerged after Xu *et al.*work, recently, Sun *et al.* [156] has identified

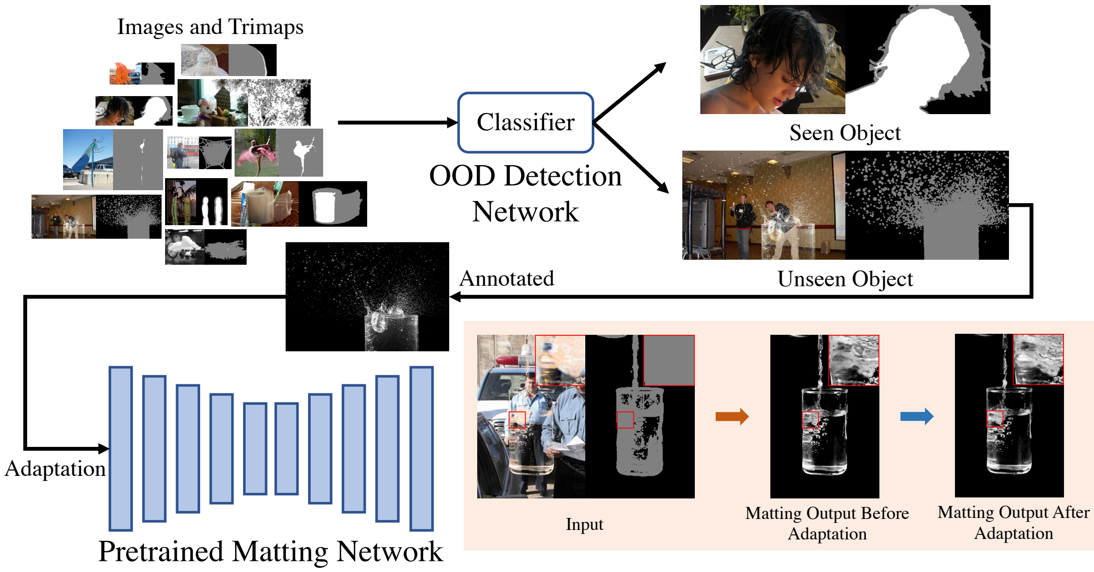

Figure 1.1: The overview of our open set matting (OSM) framework. The out-of-distribution (OOD) detection network detects unseen samples whose appearance within unknown region of trimap is unseen during training. After annotation of a few unseen samples, we conduct few-shot adaptation.

a bias issue of previous matting datasets, including AIM [177] and the Distinctions-646 dataset [125]. 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 [137], 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 application, 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.

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 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 the connectivity between samples [142]. 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, e.g., 11.58%, 42.80%, and 28.57% relative improvements in AUROC, FPR95(OUT), and Detection Error metrics compared to the state-of-art 1-D Subspaces [193]. We conduct experiments and analysis to validate the effectiveness of our few shot learning matting module.

To summarize, our contributions are as follows:

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

Chapter 2

Related Work

In this section, we cover related works about natural image matting, out-of-distribution detection, prototype learning, incremental learning, open set domain adaptation, and open set learning.

2.1 Natural Image Matting

Natural image matting is a task to estimate alpha mattes to precisely extract foregrounds from background images. Traditional image matting can be roughly categorized into sampling-based [4, 31, 52, 61, 146, 166, 167] and propagation-based [28, 57, 70, 80–82, 153, 175] approaches that usually require trimap as additional input. Sampling-based approaches [4,31,52,61,146,166,167] sample colors from foreground and background pixels for each unknown region defined by trimap and then leverage quantitative metrics to select the best foreground-background color pair for computing alpha values. Propagation-based approaches [28, 57, 80–82, 153] propagate alpha values from known pixels to unknown ones based on some similarity measurements.

Recently, deep learning has shown its prominent performance on matting task. Learning-based image matting can be divided into three primary categories, i.e., trimap-needed [29, 35, 86, 94, 102, 104, 121, 156, 157, 177, 189, 191], background-required [89, 124, 143, 179], and only-image [75, 83–85, 125, 170, 200] input, i.e., trimap-free.

For background-required matting, background matting [143], with an image and its corresponding background and soft segmentation as input, obtains appealing estimation but is not robust to images with shadow or under complex light conditions. Lin *et al.* [89] introduce a real-time, high-resolution background matting technique along with two large-scale video and image matting datasets, i.e., VideoMatte240K and PhotoMatte13K/85.

For trimap-needed matting, Cho *et al.* [29] first introduce deep neural networks into image matting task. Xu *et al.* [177] propose a deep neural matting network as well as a comprehensive matting dataset that has significantly promoted research progress. Lutz *et al.* [104] explore matting task with a generative adversarial framework. Then, appealing matting results are achieved by Lu *et al.* [102] and Tang *et al.* [157]. Subsequently, GCA-Matting [86], a matting network with guided contextual attention, not only simulates information flow of affinity-based methods but also models matting in a view of image inpainting. HDMatt [189] attempts to tackle high-resolution matting problem by introducing three non-local attentions to propagate each trimap region of context patches into the corresponding region of query patches. Semantic Image Matting (SIM) [156] leverages semantic information to boost matting performance by introducing Semantic Image Matting Dataset (SIMD), which divides matting data into 20 categories. MGMatting [191] takes a general coarse mask as guidance and leverages a progressive refinement network design to provide self-guidance and progressively refine uncertain regions in the decoding process. MatteFormer [121] proposes the first transformer-based image matting approach that uses trimap as prior-token to participate in the self-attention mechanism of each prior-attentive swin transformer block.

Since trimap-needed matting methods usually require high-quality handmade trimap whose production process is time-consuming and background-required matting needs an inconvenient background acquisition process, a few works attempt to take only images as input and produce alpha mattes. Trimap-free matting works can be distinguished into two families, including single-category trimap-free matting, such as human/animal matting [27, 83, 84, 93], and general trimap-free matting, also known as automatic image matting [85, 125, 191, 200]. For human/animal matting, Chen *et al.* [27] propose an automatic human matting algorithm, Semantic Human Matting (SHM), to learn semantic information and high quality details from data with a novel fusion strategy. Liu *et al.* [93] refine coarse masks sequentially by coupling coarse annotated data with fined one to improve end-to-end semantic human matting without trimap as extra input. Li *et al.* [84] propose a Glance and Focus Matting network (GFM) that employs a shared encoder and two separate decoders to learn high-level semantic segmentation and low-level details collaboratively with AM-2k and BG-20k datasets. They also investigate the domain gap issue between synthetic images and real-world images and design a composition route RSSN to improve the model generalization ability. Li *et al.* [83] introduce P3M-10k benchmark, the first large-scale anonymized benchmark for Privacy-Preserving Portrait Matting, and a corresponding strong baseline P3M-Net based on GFM. Sun *et al.* [154] propose Semantics-Adding Flaw-Erasing network (SAFE-Net) to first compute an initial alpha matte, detect its errors by a flaw detector, and then correct it using a refinement process. They also construct a large human matting dataset containing 4,729 unique foregrounds with fine annotations. For the general trimap-free matting methods, Zhang *et al.* [200] design two convolutional decoder branches for foreground and background classifications, and then a fusion branch to integrate two classification results for trimap-free matting. Qiao *et al.* [125] introduce the Distinction-646 trimap-free matting dataset and HAttMatting by using channel/spatial-wise attention to filter out noise from hierarchical appearance cues and boost alpha mattes. Yu *et al.* [191] use general coarse masks as guidance for matting but could suffer from performance degradation due to inaccurate masks. Li *et al.* [85] propose AIM-Net by investigating the possibility of extending to images with salient transparent/meticulous foregrounds and non-salient foregrounds with unified semantic representation. However, the shared encoder of AIM-Net might be unable to fully represent semantic/matting information, resulting in artifacts of "hard fusion" between results from semantic and matting decoders. And since their proposed Automatic Image Matting-500 benchmark mostly contains clean/blurred backgrounds without complicated multiply objects that conforms with real-world photography but simplifies the problem, their approach could degrade when encountering images with clear/complicated backgrounds.

Different from previous matting works, we notice that past matting works only focus on closed-set datasets, where training and test data are assumed to be drawn from a similar distribution, and then can be problematic in real application. Therefore, we introduce an open set matting framework that can detect OOD to-be-matted objects, deliver them to human for post-processing, and then obtain better results on OOD to-be-matted objects.

2.2 Out-of-Distribution (OOD) Detection

Out-of-distribution (OOD) detection aims to detect testing samples drawn from a different distribution compared to training samples. OOD detection can be roughly categorized into two domains, i.e., uncertainty estimation-based methods and generative model-based methods.

For uncertainty estimation-based OOD detection approaches, the maximum softmax probability (MSP) [64] serves as a baseline for uncertainty estimation. Hendrycks *et al.* [63] 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 a wrong prediction with high confidence in that DNNs are usually poorly calibrated [59]. Therefore, there are many works that aim to achieve better uncertainty estimation. For example, Guo *et al.* [59] evaluate various post-processing calibration methods and provide a temperature scaling solution at calibrating predictions. Monte Carlo dropout (MC-dropout) [49] and ensembles [79] approaches leverage approximate Bayesian inference to better estimate uncertainty. Maddox *et al.* [105] propose Stochastic Weight Averaging-Gaussian (SWAG), which fits a Gaussian using Stochastic

Weight Averaging as the first moment and derives a low rank plus diagonal covariance from SGD iterations to form an approximate posterior distribution over neural network weights with sampling based Bayesian model averaging. Thulasidasan *et al.* [158] find that DNNs trained with mixup [197] or label smoothing in mixup training are significantly better calibrated. Furthermore, Zaeemzadeh *et al.* [193] 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).

For reconstruction-based OOD detection approaches, its core idea is that a reconstruction network trained on ID data usually produces smaller reconstruction error for ID data than that for OOD data [40]. The auto-encoder (AE) [3,11], variational auto-encoder (VAE) [9], GAN [194], U-Net [95], and Restricted Boltzman Machine neural network (RBM) [34] are popular backbones that can be used for this method. Zhou *et al.* [202] introduce semantic reconstruction, data certainty decomposition, and normalized L2 distance to improve auto-encoder based methods without any extra data, complicated structure, time-consuming pipeline, and degrading the classification accuracy of seen classes. Yang *et al.* [182] propose an OOD detection framework, MOODCat, for image classifiers that can naturally learn the semantic information of in-distribution data. MOODCat randomly masks out a portion of an input image, leverages conditional GAN to synthesize a new image with the masked image as input according to the classification result, and then adopts the semantic difference between the original image and synthesized one for OOD detection.

2.3 Prototype Learning

Prototype learning is a deep learning counterpart of traditional nearest neighbor classification and Learning Vector Quantization (LVQ) [77] 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 [45, 162], OOD/anomaly detection [19, 26, 180, 181], few-shot learning [117, 152, 169], and person re-identification [185]. 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 where 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.

Prototype learning loss function can be divided into classification-based loss functions, e.g., large-margin softmax loss [97] and center loss [171], and distance-based loss functions, e.g., contrastive loss [30,155], triplet loss [140,141], and tuplet margin loss [187]. For classification-based loss functions, center loss minimizes the distance between each example and its defined class center to form a class-dependent constraint while the large-margin softmax loss [97] defines more rigorous boundary for correct classification training. Then, the large-margin softmax loss has been further improved by feature normalization [99, 126, 164, 201] and weight normalization [98]. Furthermore, different types of margin, such as additive cosine margin [163, 165] and additive angular margin [39], have also been introduced to improve the large-margin softmax loss. For distance-based loss functions, triplet loss is formed by triplets, each of which consists of an anchor, a positive, and a negative sample, and is designed to make the distance between the anchor and positive sample smaller than the distance between the anchor and negative sample. Tuplet margin loss [187] disentangles the norm and direction of feature embedding to implicitly up-weight hard samples and down-weight easy samples with a slack margin in angular space and mitigates over-fitting on the hardest sample. This loss also reduces the influence of class-dependent information to improve the generalization ability.

2.4 Incremental Learning

Deep learning have exhibited its impressive performance on individual learning task, but are unable to scale its behavior over time and could suffer from catastrophic forgetting when new data becomes available. To address this issue, incremental learning investigates the task of gradually extending existing knowledge to an infinite stream of data. Incremental learning methods can be distinguished into three families [36], i.e., replay methods, regularization-based methods, and parameter isolation methods.

Replay methods keep samples or generate pseudo-samples using a generative model and then replay previous task samples while adapting into a new task to alleviate forgetting. These samples are either re-used as model inputs for rehearsal [25, 37, 67, 127, 130, 147], or involved in constraining the optimization of the new task loss to prevent interfering with previous tasks [8, 24, 101].

Different from replay methods, regularization-based methods avoid storing previous task samples and introduce an extra regularization term into the loss function which consolidates previous knowledge while learning new data. These methods can be further categorized into data-focused and prior-focused methods. The basic idea of data-focused methods is knowledge distillation from a pretrained model on the previous task to the model on the new data [73, 88, 151, 199]. For example, Learning without Forgetting (LwF) [88] trains on new data by optimizing the new task accuracy and response preservation of the existing tasks from the original network. Prior-focused methods [1, 5, 7, 76, 76, 116, 195, 196] estimate the importance of model parameters for the previous task and consider it as prior to constrain the change of important parameters during the next-stage adaptation. For instance, Elastic Weight Consolidation (EWC) [76] is the first well-established work. Zenke *et al.* [195] simplifies EWC to estimate importance weights online during task training.

Parameter isolation methods leverage different model parameters for each task to prevent catastrophic forgetting. For example, Rusu *et al.* [134] and Xu *et al.* [176] keep growing new branches for new tasks while freezing previous task parameters. Aljundi *et al.* [6] allocate model copy for each task. Alternatively, some approaches [47, 106, 144] keep the network architecture static but allocates fixed parts of the network architecture to each task. Fernando *et al.* [47] introduce embedded agents to a neural network, whose task is to discover which parts of the network should be re-used for new tasks, denoted as PathNet. Mallya *et al.* [106] propose an approach called PackNet. PackNet first uses weight-based pruning techniques to free up redundant parameters across all layers of a deep neural network trained for previous tasks with minimal accuracy degradation. Then, it modifies the freed up parameters while preserving the surviving parameters for learning a new task. Serra *et al.* [144] introduce a hard attention mechanism that preserves previous tasks' information but does not affect the current task's learning. Each task learns its corresponding hard attention mask concurrently with previous task masks applied to constrain the learning.

2.5 Open Set Domain Adaptation

Open set domain adaptation (OSDA) attempts to align distributions of shared classes between source and target domains while also perceiving unseen classes in the absence of target domain labels. Busto *et al.* [120] consider open set domain adaptation as an assignment task. Saito *et al.* [136] leverage adversarial training for an open set domain adaptation scenario where unseen samples only exist in the target domain. Specifically, they train a generator to extract features, which distinguishes unseen target samples from seen target samples, with a classifier building the boundary between seen and unseen classes. Liu *et al.* [91] consider the openness of the target domain and introduce the Separate to Adapt (STA) method that divides the training procedure into a seen/unseen separation step and a weighted adversarial adaptation step to progressively separate seen and unseen samples and align features of seen classes between source and target domains. Feng *et al.* [46] introduce Semantic Categorical Alignment and Semantic Contrastive Mapping to exploit the semantic structure of open set data and enlarge margins across seen classes and margins between seen and unseen classes. Pan *et al.* [119] introduce category-agnostic clustering in the target domain into Self-Ensembling. Further, instead of grouping unseen samples as one generic class that might lead to suboptimal solution, they exploit the inherent data structure by clustering in the target domain to learn domain-invariant representations of seen classes and discriminative representations among seen and unseen classes in the target domain. Xu *et al.* [178] propose the Joint Partial Optimal Transport (JPOT) approach that uses the joint discriminative prototypical compactness loss to obtain intra-class compactness and inter-class separability and estimate the mean and variance of the unseen class. Luo *et al.* [103] propose a Progressive Graph Learning (PGL) framework with adversarial learning to achieve a tighter upper bound of the target error. Bucci *et al.* [15] introduce the self-supervised task of predicting image rotation into open set domain adaptation to achieve a competitive performance. Jing *et al.* [71] propose a balanced OSDA approach that finds the trade-off between the accuracy of identifying the unseen samples and the classification ability of the seen samples by projecting the feature into a hyper-spherical latent space and bounding the centroid deviation angle. Saito *et al.* [135] introduce a one-vs-all classifier for each class in the source domain to learn both inter-class and intra-class distances for the separation of seen and unseen samples and then apply an open set classifier in the target domain for OSDA. Jang *et al.* [69] propose Unknown-Aware Domain Adversarial Learning (UADAL) that aligns the source and the target-seen distribution and simultaneously segregates the target-unseen distribution during the feature alignment. UADAL not only learns the target-unknown feature space but also theoretically guarantees alignment and segregation. Liu *et al.* [100] propose a novel prototype-based shared-dummy classifier (PSDC) model for OSDA, which consists of three training steps, i.e., open set recognition, weighted alignment, and pseudo-unknown learning. First, PSDC is applied to the source domain in order to estimate class prototypes and place the dummy prototypes near class decision boundaries. Second, PSDC estimates the novel class distributions using the generated open-set samples by pushing decision boundaries tighter. Third, the weighted alignment module adapts the trained model to the target domain. Fourth, the pseudo-unknown learning module pushes the selected pseudo-unknown target instances away from the estimated prototypes by maximizing entropy. Bucci *et al.* [14] propose a distance-based hyper-spherical classification method, denoted as HyMOS, for multi-source open set domain adaptation. Liu *et al.* [92] introduce a novel Unknown-Oriented Learning (UOL) framework with a multi-unknown detector and graph learning for OSDA, which consists of three stages, including true unknown excavation, false unknown suppression, and known alignment, in order to explore the rich semantic information and intra-class variation of unknown classes. Yu *et al.* [190] introduce a new problem setting in OSDA, that is, few-shot OSDA. They propose a self-labeling framework that leverages prototypical contrastive learning and maximizes the mutual information between labels and input data to classify seen and unseen classes in source and target domains.

2.6 Open Set Learning

Traditional supervised learning assumes a situation where training and testing samples follow the same distribution while open set learning is a more challenging and realistic setting where there are unseen testing samples. Except well-studied open set recognition (OSR) [137], there are open world segmentation [19] and open set action recognition [10]. To the best of our knowledge, this is the first work that tackles matting problem in an open-set perspective. Here we focus on well-investigated OSR. OSR requires classifiers to not only accurately classify seen class but also effectively deal with unseen classes. In this section, we will first cover some basic notations and definitions related to OSR and then introduce OSR by categorization.

2.6.1 Basic Notations and Definitions

Here we briefly review the OSR formulation by Scheirer *et al.* [137]. The space far from seen data is usually considered as open space $\mathcal{O}$. Labeling any sample in open space as an arbitrary seen class incurs risk, which is called open space risk $R_{\mathcal{O}}$. Scheirer *et al.* [137] formulate $R_{\mathcal{O}}$ as the relative measure of open space $\mathcal{O}$ compared to the overall measure space S_o,

$$R_{\mathcal{O}}(f) = \frac{\int_{\mathcal{O}} f(x)dx}{\int_{S_o} f(x)dx}, \tag{2.1}$$

where f denotes the measurable recognition function. $f(x) = 1$ means recognizing sample x as one seen class, otherwise $f(x) = 0$. According to Eq. (2.1), the more samples in open space is considered as seen classes, the larger $R_{\mathcal{O}}$ is.

Further, Scheirer *et al.* [137] also introduce the concept of openness, which can evaluate the difficulty level of OSR:

Definition 1 (The Openness) *Let C_{target}, C_{train}, and C_{test} respectively represent the set of seen classes during testing, the set of seen classes in training, and the set of seen and unseen classes used during testing. Then, the openness of the corresponding recognition task O is:*

$$O = 1 - \sqrt{\frac{2 \times |C_{\text{train}|}}{|C_{\text{target}}| + |C_{\text{test}}|}}, \tag{2.2}$$

where $|\cdot|$ denotes the number of classes in the corresponding set.

Larger openness indicates more open and challenging problems while zero openness means closed-set problem.

After introducing the concepts of open space risk and openness, the definition of OSR problem can be given as follows:

Definition 2 (The Open Set Recognition Problem [137]) *Considering training data V, an open space risk $R_{\mathcal{O}}$, and a data error measure D, open set recognition is to find a measurable recognition function $f \in \mathcal{H}$, where $f(x) > 0$ implies correct recognition, and f is defined by minimizing the open set risk:*

$$\arg\min_{f \in \mathcal{H}} \left\{ R_{\mathcal{O}}(f) + \lambda_r D(f(V)) \right\}, \tag{2.3}$$

where λ_r is a regularization constant.

The open set recognition is to minimize the open set risk, which balances the tradeoff between the open space risk and empirical data error terms, over the space of all possible inputs.

2.6.2 Open Set Recognition (OSR)

OSR can be categorized into two domains [54], including traditional and deep learning based methods.

Traditional OSR methods

There are many attempts [12,20–23,44,68,107,113,133,137–139,159,161,198] that extends traditional machine learning methods, e.g., Support Vector Machine (SVM) [32], sparse representation, and nearest neighbor, to the OSR setting.

(1) SVM based: Based on SVM, Scheirer *et al.* [137] propose 1-vs-Set machine that adds another hyperplane paralleled with the separating hyperplane obtained by SVM and initializes the planes to contain all positive training data. Then, the open space risk is

$$R_{\mathcal{O}} = \frac{\delta_\Omega - \delta_A}{\delta^+} + \frac{\delta^+}{\delta_\Omega - \delta_A} + p_A\omega_A + p_\Omega\omega_\Omega, \tag{2.4}$$

where δ_A and δ_Ω denote marginal distances of the near and far planes respectively, δ^+ is the separation needed to account for all positive data, $\frac{\delta_\Omega - \delta_A}{\delta^+}$ is the expansion of plane distance that serves as the overgeneralization risk, and $\frac{\delta^+}{\delta_\Omega - \delta_A}$ is the overspecialization risk. Moreover, user-specified parameters p_A and p_Ω weight the importance between the margin space ω_A around the near plane and ω_Ω around the far plane. In this case, a testing sample that appears within these two hyperplanes would be labeled as positive; otherwise, it would be considered as non-target class or rejected, depending on which side of the hyperplane it resides. Cevikalp *et al.* [20, 23] propose the best fitting hyperplanes approach that makes each hyperplane close to the samples of one of the classes and far from the other class samples. Scheirer *et al.* [138] incorporate non-linear classifiers with open space risk limiting classification in a multi-class setting. They introduce a compact abating probability (CAP) model, where the probability of class membership decreases as points move away from seen data to open space, and a Weibull-calibrated SVM (W-SVM) method that is integrated with the extreme value theory (EVT) [50]. Jain *et al.* [68] leverage EVT to model positive training samples at the decision boundary and introduce P_I-SVM that estimates the unnormalized posterior probability of class inclusion. Scherreik *et al.* [139] introduce probabilistic open set SVM (POS-SVM) that can empirically determine rejection threshold for each seen class. Cevikalp *et al.* [21, 22] introduce a family of quasi-linear "polyhedral conic" discriminants that define acceptance regions for positive seen classes and help discriminate positive samples from negative ones.

(2) Sparse Representation based: The sparse representation-based classifier (SRC) [123, 132, 173, 174] has been widely used in computer vision and image processing fields, which makes the prediction by seeking sparsest representations for testing samples with learned redundant dictionaries. In order to adapt SRC into OSR, Zhang *et al.* [198] introduce the sparse representation based open set recognition method, SROSR. SROSR models the tail part of the matched and sum of nonmatched reconstruction error distributions using Extreme Value Theory (EVT) to simplify OSR into hypothesis testing problems.

(3) Distance-based: Bendale *et al.* [12] extend Nearest Class Mean (NCM) classifier [108, 129] to OSR and introduce a Nearest Non-Outlier (NNO) algorithm. The classification of NNO is based on the distance between the testing sample and the means of seen classes and rejecting a sample if all classifiers reject it. And NNO can dynamically incorporate new classes by using manually labeled data. Júnior *et al.* [107] introduce an open set Nearest Neighbor classifier (OSNN). OSNN leverages a Nearest Neighbor Distance Ratio (NNDR) technique that thresholds the ratio of similarity scores between a test sample and its two most similar classes.

(4) Margin Distribution based: Since there is few OSR work taking distribution information into consideration, Rudd *et al.* [133] propose theoretically sound Extreme Value Machine (EVM) based on margin distribution [2, 51, 122, 128] and Extreme Value Theory

(EVT). EVM can perform the nonlinear kernel-free variable bandwidth OOD detection with incremental learning and has been applied to open set face recognition [58] and open set intrusion detection [65]. However, EVM suffers in situations where the geometries of seen and unseen classes are different. In order to address this limitation, Vignotto *et al.* [161] propose GPD and GEV classifiers relying on approximations from EVT.

Deep learning based OSR methods

(1) Discriminative methods: Recently, Deep Neural Networks (DNNs) have shown impressive performance in classification. Since DNNs usually leverage a SoftMax cross-entropy loss, which involves normalization and has a closed-set nature, DNNs might make the prediction wrongly with high confidence when encountering unseen classes [56, 115]. To this end, there are various works attempting to solve this problem [13, 16, 17, 41, 60, 74, 118, 131, 148, 149, 160, 186]. Bendale *et al.* [13] replace SoftMax layer with OpenMax layer as a solution for open set deep networks. Specifically, after training a deep neural network with the SoftMax cross-entropy loss, they consider the network values from the penultimate layer as the activation vector and represent each class as a mean activation vector (MAV) that is calculated by correctly classified training samples. Then, they calculate the distances between training samples and their corresponding MAVs to fit the separate Weibull distribution for each class. Further, they redistribute the activation vector's values by the Weibull distribution fitting score and use it for final classification and unknown/unseen classes rejection. Hassen *et al.* [60] introduce a neural representation which maintains inter-class separability and intra-class compactness. Venkataram [160] follow OpenMax to explore open set text classification. Shu *et al.* [148] propose a Deep Open Classification (DOC) method with a 1-vs-rest final layer containing a sigmoid function for each seen class. Kardan *et al.* [74] introduce the competitive overcomplete output layer (COOL) neural network to tackle the overgeneralization issue of neural networks over regions far from the training data. Cardoso *et al.* [16, 17] illustrate how to use WiSARD, a weightless neural network model, for open set recognition, which is based on an elaborate distance-like computation a weightless neural network provides. Dhamija *et al.* [41] introduce novel Entropic Open-Set and Objectosphere losses for OSR. Yoshihashi *et al.* [186] present the Classification-Reconstruction learning algorithm for open set recognition (CROSR) that leverages latent representations for reconstruction and leads to robust OOD detection without hurting the classification accuracy of seen classes. Oza *et al.* [118] introduce class conditioned auto-encoders with novel training and testing methods for OSR. Shu *et al.* [149] focus on discovering hidden unseen classes of rejected samples and introduce a joint open classification model with a sub-model distinguishing whether a pair of samples resides in the same class or not. This sub-model can be considered as a distance function for clustering to discover hidden classes of rejected samples.

(2) Generative methods: Since generative adversarial networks (GAN) [55] have shown impressive performance, there are some OSR works [53, 72, 112, 183, 192] leveraging unseen samples generated by GAN. Ge *et al.* [53] extend OpenMax by employing generative adversarial networks for unseen class image synthesis, denoted as Generative OpenMax (G-OpenMax). G-OpenMax can provide explicit modelling and decision score for unseen classes. Neal *et al.* [112] introduce a novel dataset augmentation technique, called counterfactual image generation, which generates synthetic samples that are close to training samples but do not belong to any training category, for OSR. Jo *et al.* [72] adopt GAN to generate fake data as unseen data to further enhance the robustness of classifiers for OSR by modelling noisy distribution on the classifier feature space using the proposed marginal denoising auto-encoder. Yu *et al.* [192] propose an adversarial sample generation (ASG) framework for OSR, which generates positive and negative samples of seen classes in an unsupervised manner. Yang *et al.* [183] propose OpenGAN, where the generator produces fake samples and the redesigned discriminator outputs multiple classes together with an unseen class, for OSR.

Chapter 3

Approach

In this section, we first define our problem setup and then introduce our OOD detection network and incremental few-shot learning matting network.

3.1 Problem Setup

In Figure 1.1, 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 $C_{\text{in}+t} = C_{\text{in}} \cup \{C_{\text{out},1}, C_{\text{out},2}, ..., C_{\text{out},t}\}$, $t \in \{1, 2, ..., K\}$.

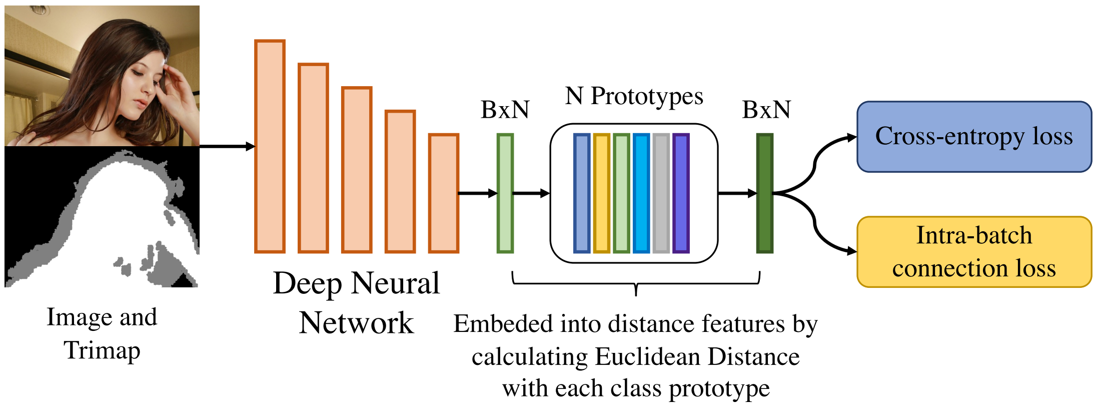

Figure 3.1: Our OOD Detection Network (OOD-DN). Our OOD-DN leverages prototype learning with intra-batch connection to be unseen-aware and generate informative logit features.

3.2 OOD Detection Network (OOD-DN)

Figure 3.1 shows our OOD detection network that can be disentangled into a feature extractor and a discriminant function. The ResNet-50 [62] 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 [19, 109, 180]. 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.

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

$$d_i = -||f(X;\theta_f) - p_i||_2^2. \tag{3.1}$$

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

$$\mathcal{L}_{\text{CE}} = -\log\left(\frac{\exp(d_y)}{\sum_{i=1}^{N}\exp(d_i)}\right), \tag{3.2}$$

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.

3.2.1 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

$$\mathcal{L}_{\text{IBC}} = \sum_{i=1}^{N}\sum_{j=1,j<k}^{C_i} D_{\text{KL}}(D_{C_i}^{(j)} || D_{C_i}^{(k)}), \qquad D_{\text{KL}}(\boldsymbol{p}||\boldsymbol{q}) = \sum_{i=1}^{N}\boldsymbol{p_i}\log\left(\frac{\boldsymbol{p_i}}{\boldsymbol{q_i}}\right), \tag{3.3}$$

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.

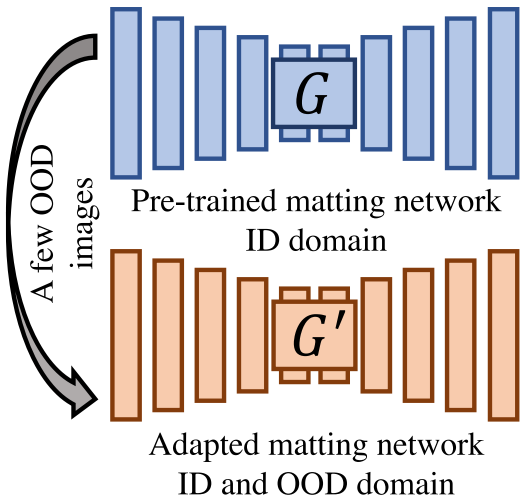

Figure 3.2: The pipeline of our incremental few-shot learning matting module (IFL-MM). After training the matting network on ID data, we extend it to OOD data with a few samples but without catastrophic forgetting.

3.2.2 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 [63]. Specifically, given input X, the anomalous score is defined as

$$S(f(X;\theta_f)) = -\max(d_i), \quad i \in \{1, 2, ..., N\}. \tag{3.4}$$

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, as shown in Figure 3.2, 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 [86] 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.

3.3.1 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) [66]; (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 [110,172]; (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 [87]. 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.

3.3.2 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) [76] is a regularization approach that aims to overcome catastrophic forgetting by constraining the model weights 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,

$$L = |\alpha - \hat{\alpha}| + \frac{\lambda}{2} \cdot \sum_i F_i \cdot (\theta_i - \theta_i^*)^2, \tag{3.5}$$

where α is the alpha matte label, θ_i^* is the optimal value from previous tasks and λ is the regularization hyperparameter. The diagonal of the Fisher information matrix F, which can be computed from first-order derivatives alone, is equivalent to the second derivative of the loss near a minimum.

Synaptic Intelligence (SI) [195] is a simplified variant of EWC. Instead of expensive Fisher information matrix computation, F_i is calculated online by integrating the loss over the weight trajectories during gradient descent,

$$\Delta L_i = \Delta\theta_i \cdot \frac{\partial L}{\partial \theta_i}, \tag{3.6}$$

$$F_i = \frac{\sum \Delta L_i}{\Delta_i^2 + \xi}, \tag{3.7}$$

where $\Delta\theta_i$ is the weight update amount during each training step, $\frac{\partial L}{\partial \theta_i}$ is the gradient, Δ_i is the total weight parameter distance, and ξ is a small constant for numerical stability.

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

$$L = |\alpha - \hat{\alpha}| + \frac{\lambda}{2} \cdot \sum_i F_i \cdot (\theta_i - \theta_i^*)^2, \tag{3.8}$$

$$F_i = \mathbf{E}\left[\left(\frac{\partial}{\partial \theta_i}\mathcal{L}(\alpha|X;\theta)\right)^2\right], \tag{3.9}$$

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

Chapter 4

Experiments

In this section, we conduct experiments on SIMD to validate the effectiveness of our OOD detection network, incremental few-shot learning matting network, and open set matting framework.

4.1 OOD Detection Network (OOD-DN)

4.1.1 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.* [145], 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 [156], we randomly composite training foregrounds with randomly selected background images from COCO [90]. For the test set, we follow Sun *et al.* [156] 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.

Methods	AUROC(IN)↑	AUPR(IN)↑	FPR95(IN)↓	AUROC(OUT)↑	AUPR(OUT)↑	FPR95(OUT)↓	DetectionError↓
MSP [64]	0.673	0.879	0.882	0.673	0.360	0.621	0.332
MaxLogit [63]	0.623	0.855	0.959	0.623	0.290	0.740	0.363
EnergyScore [96]	0.605	0.847	0.995	0.605	0.278	0.751	0.363
1-D Subspaces [193]	0.734	0.896	0.795	0.734	0.501	0.722	0.322
MMSP [19]	0.660	0.864	0.941	0.660	0.328	0.837	0.360
EDS [19]	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 4.1: OOD detection results on SIMD dataset.

λ	PL	$\mathcal{L}_{\text{CE}}$	$\mathcal{L}_{\text{IBC}}$	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 4.2: Ablation study results of our OOD detection network on SIMD dataset. PL refers to prototype learning.

4.1.2 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.

4.1.3 Implementation Details

We follow similar data processing and augmentation procedure as GCA-Matting [86] 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 $\mathcal{L}_{\text{IBC}}$, λ is set to 0.1 and m is 3.

4.1.4 Results

We compare our OOD-DN results with MSP [64], MaxLogit [63], energy score [96], 1D-subspaces[1] [193], metric-based maximum softmax probability (MMSP) [19], and Euclidean distance sum (EDS) [19]. We reproduce these methods according to their official public available code under the same training configuration as ours. In Table 4.1, 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 11.58% in AUROC(IN), 11.58% in AUROC(OUT), 42.80% in FPR95(OUT), and 28.57% in Detection Error metrics. Besides, different from 1D-subspaces that needs sample from training data after training to compute the first singular vector of each class for the later OOD detection inference, our OOD-DN does not require this time-consuming sampling procedure and can directly use the negative maximum value of logits as the anomalous score during inference.

4.1.5 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 4.2. 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

[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.

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)

4.2.1 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.

4.2.2 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 experiments 10 times and report average metrics and their corresponding standard deviations.

4.2.3 Implementation Details

In the initial training stage, we adopt the similar training strategy as GCA-Matting [86] 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, ro-

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 4.3: 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 4.4: Detailed quantitive matting results of 20 classes of SIMD dataset on SAD metric. Bolden classes are OOD classes, otherwise classes are ID classes.

tation, 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 $2e8$.

4.2.4 Results

We compare our IFL-MM with pre-trained and fine-tuned models. The Table 4.3 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

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

Table 4.5: 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).

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 4.4. 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 Figure 4.2 on 5 challenging OOD classes. Our IFL-MM can better separate target objects from background without apparent background ghost or missing foreground details and obtain overall visual improvement compared to both pre-trained and fine-tuned models.

4.2.5 Ablation Study

We conduct ablation study on our IFL-MM to investigate the effectiveness of each component. In Table 4.5, 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 Figure 4.5, 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 4.5, our test performance

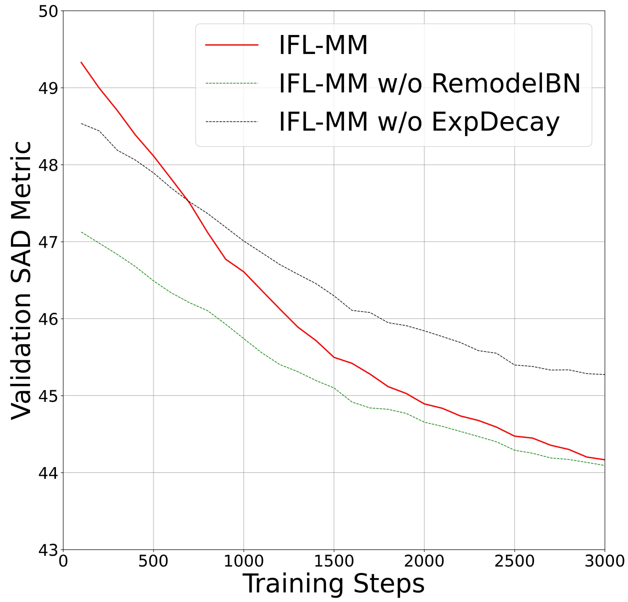

Figure 4.1: Comparison of training process among IFL-MM w/o RemodelBN, IFL-MM w/o ExpDecay, and IFL-MM on SAD metric of validation set.

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.

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 4.3. The detailed results of each class are presented in Table 4.4. 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 Figure 4.2. It is noted that our OSM significantly improves OOD

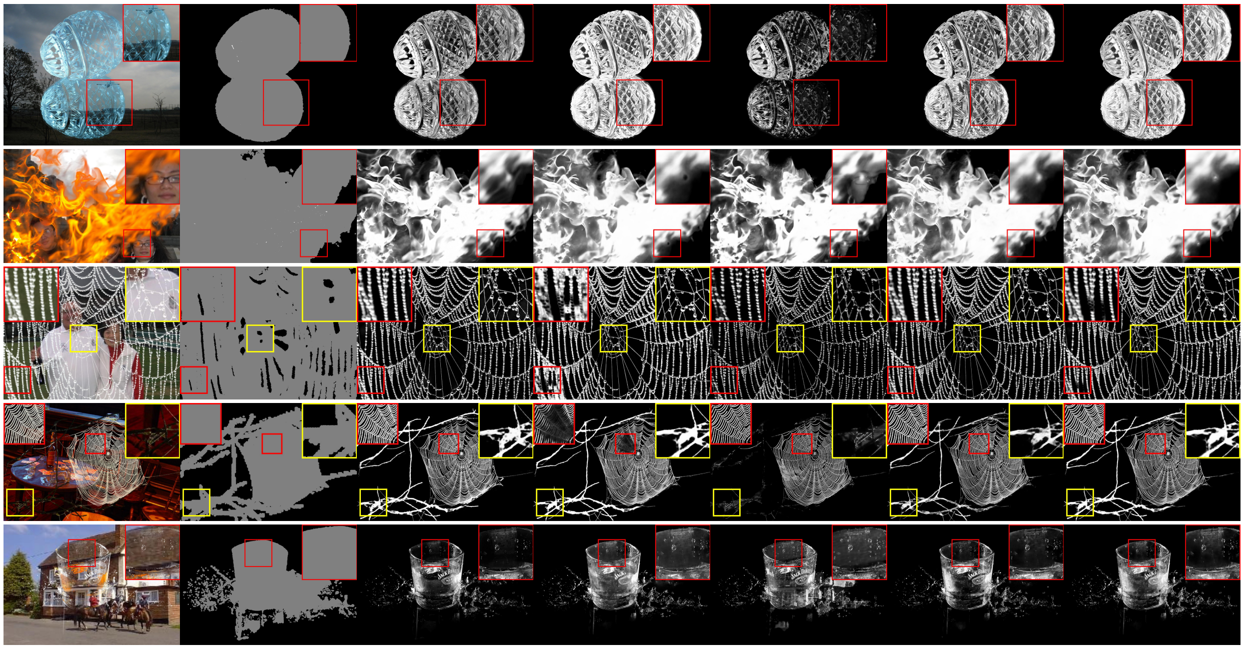

Figure 4.2: Visual comparison of matting results on 5 OOD classes of SIMD dataset. From the 1st row to the 5th row, glass_ice, fire, water_drop, spider_web, and water_spray. From left to right, image, trimap, GT, Pre-trained model, Finetune, IFL-MM (Ours), and OSM (Ours).

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 [12, 38, 42, 150] 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.

Chapter 5

Discussion

In this section, we discuss how/why we select such OOD data for our main experiments, our limitations, and possible future works.

5.1 Out-of-distribution Class Selection

Out of 20 classes in SIMD [156], we select 5 classes, i.e., glass_ice, fire, water_drop, spider_web, and water_spray, as OOD data, instead of doing leave-one-out cross validation, where all these 5 OOD classes can be considered as transparent objects. The reason behind this is as follows:

- We hope to have a similar setup as Shaban *et al.* [145]. Shaban *et al.* [145] propose a new benchmark, PASCAL-5^i, for k-shot image segmentation on PASCAL VOC 2012 [43] dataset, where they sample 5 classes out of 20 classes as test data, i.e., OOD data.
- Other OOD detection works [96, 193] also have OOD test sets with more than one OOD class, such as LSUN (10 classes) [188] and SVHN (10 classes) [114] datasets.

- Leaving more than one class out allows us to evaluate methods on a more challenging benchmark where multiple OOD classes need to be matted using a single model.

- Since natural image matting is a regression task, a pretrained matting model, no matter which data it is trained on, can be directly applied to any images. Therefore, domains involved in adaptation should be distinguished enough to make open set matting meaningful. For trimap-free setting [85], popular domain gaps natural image matting will encounter can be *human/animals* $\leftrightarrow$ *all object categories except human/animals*, *salient solid objects* $\leftrightarrow$ *salient transparent/meticulous objects*, *salient solid objects* $\leftrightarrow$ *non-salient objects*, and *salient transparent/meticulous objects* $\leftrightarrow$ *non-salient objects*. Please refer to [85] for trimap-free object categories. For trimap-needed setting, domain gaps might be narrowed down to *solid objects* $\leftrightarrow$ *transparent objects* and *human/animals* $\leftrightarrow$ *all object categories except human/animals*. Therefore, as the first step towards open set matting, we attempt to tackle *solid objects* $\rightarrow$ *transparent objects* in trimap-needed setting.

- The leave-one-out cross validation would require a huge number of experiments (twenty per method in our case) which is computationally prohibitive considering the number of hyperparameters we have.

5.2 Limitations

- Our OOD-DN cannot maintain excellent performance on both classification of ID classes and OOD detection at the same time.

- According to our experiments, our IFL-MM is sensitive to OOD-ID split of dataset. More advanced incremental few shot learning approaches can be adapted to improve performance.

- We does not investigate how to increase classes in OOD-DN and incorporate information of OOD-DN with IFL-MM.
- We tackle open set matting in trimap-needed aspect instead of more meaningful trimap-free aspect.
- We tackle the situation where there is only one or more than one single-type object in each image instead of multiply-type objects.

5.3 Future Works

5.3.1 Dataset

The SIMD dataset [156] consists of 20 classes with 726 training foregrounds and 89 testing foregrounds. The full training (resp. test) sets are synthetic data by compositing foregrounds with background images from COCO [90] (resp. PASCAL VOC [43]) dataset. SIMD can be considered as a representative trimap-needed open set matting benchmark. For trimap-free setting, one possible benchmark that can be used in open set matting task is Automatic Image Matting-500 Benchmark (AIM-500) benchmark [85]. AIM-500 is a high-resolution natural image matting test set, including 424 salient opaque (SO), 43 salient transparent/meticulous (STM), and 33 non-salient (NS) images. To evaluate on it, the training set is the combination of DUTS [168] and the synthetic data by compositing foregrounds of Adobe Image Matting [177], Distinction-646, and AM-2k [84] with the high-resolution BG-20k [84] background dataset. There are SO, STM, and NS type labels for synthetic training data and we consider the DUTS dataset as SO type.

As mentioned above, it is noted that, for now, there are not many matting benchmark with category labels and our benchmark option is limited. Therefore, we look forward to more matting benchmarks enabling us to tackle open set matting problem.

5.3.2 Single-type Object vs. Multiply-type Objects

As mentioned in Section 5.3.1, in test sets of SIMD and AIM-500, there is only one or more than one single-type object in each image. Under this circumstance, we can only detect if one image is OOD or not. However, for one image, if there are ID and OOD objects mixed in trimap-free setting or there are ID and OOD objects within unknown regions of trimaps in trimap-needed setting, our OOD-DN cannot handle. To this end, we have to conduct a similar experimental setup as anomaly segmentation [19] or OOD detection in multi-label setting [63] instead of multi-class, which depends on the specific problem setup. Sequentially, the incremental few-shot learning part should also be reformulated to fit the problem setup. We are encouraged to find that there are so many possibilities in open set matting task and delighted to make the first step towards this direction.

5.3.3 Trimap-needed Matting vs. Trimap-free Matting

As mentioned in Section 5.3.1, there are two possible settings in open set matting task, including trimap-needed and trimap-free. Compared to trimap-free setting, trimap-needed setting can produce superior matting results but is a bit less exciting because trimaps serve as prior information and can assist matting greatly, resulting in the situation where few-shot adaptation shows much more significant improvement in metrics than in visual results. However, open set matting in trimap-free setting would become more meaningful because of the following points:

- Trimap-free matting still cannot achieve excellent performance among comprehensive object categories. It will be exciting to improve its performance by incremental few-shot learning and enlarge its knowledge base.

- Unlike trimap-needed matting, trimap-free matting might fail easily in real application. It will be beneficial to continue updating trimap-free matting algorithm by collecting/annotating real-world OOD data and conducting fast adaptation for practical usage.

- Trimap-free matting has various kinds of OOD-ID splits, which means more possibilities for future research.

We look forward to open set matting work in trimap-free setting and believe that it will greatly contribute to real-world application of trimap-free matting.

5.3.4 Open World Matting

As discussed in Section 4.3, we validate the effectiveness of our open set matting framework; however, our OOD-DN is not incremental and does not scale gracefully with unseen classes, which is the difference between open set matting and open world matting. Although there is no open world matting work, there exist some works about open world recognition (OWR). Some OWR studies [18, 33, 48, 78, 111, 184] mainly focus on how to incorporate new classes instead of recognizing unseen classes while Bendale *et al.* [12] introduce four steps for OWR, including detecting OOD classes, choosing which samples to annotate, annotating these samples, and enlarging the classifier. Therefore, one possible future work of open set matting is to scale flexibly with the increasing number of classes for open world matting.

Chapter 6

Conclusions

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 metric-based prototype learning to embed samples into the prototype space and applies our intra-batch connection loss on this space 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 OOD-DN suppresses other state-of-the-art OOD detection methods significantly and improves AUROC, FPR95(OUT), and Detection Error metrics by 11.58%, 42.80%, and 28.57% compared to the state-of-art 1-D Subspaces [193]. Our IFL-MM takes advantage of the importance of weights of matting network trained on ID data for weight regularization and remodels the statistics of Batch Normalization with exponential learning rate decay schedule to effectively prevent catastrophic forgetting and over-fitting.

Chapter 7

Appendix

Methods	AUROC(IN)↑	AUPR(IN)↑	FPR95(IN)↓	AUROC(OUT)↑	AUPR(OUT)↑	FPR95(OUT)↓	DetectionError↓
MSP [64]	0.550	0.888	0.950	0.550	0.163	0.860	0.371
MaxLogit [63]	0.678	0.910	0.843	0.678	0.314	0.877	0.324
EnergyScore [96]	0.693	0.913	0.743	0.693	0.342	0.877	0.319
1-D Subspaces [193]	0.748	0.939	**0.664**	0.748	**0.420**	0.689	0.306
MMSP [19]	0.607	0.912	0.986	0.607	0.176	0.656	0.344
EDS [19]	0.436	0.825	1.000	0.436	0.149	1.000	0.500
OOD-DN (Ours)	**0.758**	**0.949**	0.800	**0.758**	0.337	**0.513**	**0.274**

Table 7.1: Additional OOD detection results on SIMD dataset.

We present another OOD-ID split setting of SIMD dataset where we consider 5 classes, i.e., lace, silk, net, spider_web, plastic_bag, out of 20 classes as OOD data. The comparison of our OOD-DN and other OOD detection methods is shown in Table 7.1. The results show that our OOD-DN is better than other state-of-the-art methods on most of evaluation metrics.

Bibliography

[1] AHN, H., CHA, S., LEE, D., AND MOON, T. Uncertainty-based continual learning with adaptive regularization. *Advances in neural information processing systems 32* (2019).

[2] AIOLLI, F., DA SAN MARTINO, G., AND SPERDUTI, A. A kernel method for the optimization of the margin distribution. In *Artificial Neural Networks-ICANN 2008: 18th International Conference, Prague, Czech Republic, September 3-6, 2008, Proceedings, Part I 18* (2008), Springer, pp. 305–314.

[3] AKCAY, S., ATAPOUR-ABARGHOUEI, A., AND BRECKON, T. P. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian conference on computer vision* (2018), Springer, pp. 622–637.

[4] AKSOY, Y., OZAN AYDIN, T., AND POLLEFEYS, M. Designing effective inter-pixel information flow for natural image matting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 29–37.

[5] ALJUNDI, R., BABILONI, F., ELHOSEINY, M., ROHRBACH, M., AND TUYTELAARS, T. Memory aware synapses: Learning what (not) to forget. In *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 139–154.

[6] ALJUNDI, R., CHAKRAVARTY, P., AND TUYTELAARS, T. Expert gate: Lifelong learning with a network of experts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 3366–3375.

[7] ALJUNDI, R., KELCHTERMANS, K., AND TUYTELAARS, T. Task-free continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 11254–11263.

[8] ALJUNDI, R., LIN, M., GOUJAUD, B., AND BENGIO, Y. Online continual learning with no task boundaries. *arXiv preprint arXiv:1903.08671 3* (2019).

[9] AN, J., AND CHO, S. Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE 2*, 1 (2015), 1–18.

[10] BAO, W., YU, Q., AND KONG, Y. Evidential deep learning for open set action recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 13349–13358.

[11] BAUR, C., WIESTLER, B., ALBARQOUNI, S., AND NAVAB, N. Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. In *International MICCAI brainlesion workshop* (2018), Springer, pp. 161–169.

[12] BENDALE, A., AND BOULT, T. Towards open world recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 1893–1902.

[13] BENDALE, A., AND BOULT, T. E. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 1563–1572.

[14] BUCCI, S., BORLINO, F. C., CAPUTO, B., AND TOMMASI, T. Distance-based hyperspherical classification for multi-source open-set domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2022), pp. 1119–1128.

[15] BUCCI, S., LOGHMANI, M. R., AND TOMMASI, T. On the effectiveness of image rotation for open set domain adaptation. In *European conference on computer vision* (2020), Springer, pp. 422–438.

[16] Cardoso, D. O., França, F., and Gama, J. A bounded neural network for open set recognition. In *2015 International Joint Conference on Neural Networks (IJCNN)* (2015), IEEE, pp. 1–7.

[17] Cardoso, D. O., Gama, J., and França, F. M. Weightless neural networks for open set recognition. *Machine Learning 106*, 9-10 (2017), 1547–1567.

[18] Cauwenberghs, G., and Poggio, T. Incremental and decremental support vector machine learning. *Advances in neural information processing systems 13* (2000).

[19] Cen, J., Yun, P., Cai, J., Wang, M. Y., and Liu, M. Deep metric learning for open world semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 15333–15342.

[20] Cevikalp, H. Best fitting hyperplanes for classification. *IEEE transactions on pattern analysis and machine intelligence 39*, 6 (2016), 1076–1088.

[21] Cevikalp, H., and Serhan Yavuz, H. Fast and accurate face recognition with image sets. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (2017), pp. 1564–1572.

[22] Cevikalp, H., and Triggs, B. Polyhedral conic classifiers for visual object detection and classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 261–269.

[23] Cevikalp, H., Triggs, B., and Franc, V. Face and landmark detection by using cascade of classifiers. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)* (2013), IEEE, pp. 1–7.

[24] Chaudhry, A., Ranzato, M., Rohrbach, M., and Elhoseiny, M. Efficient lifelong learning with a-gem. *arXiv preprint arXiv:1812.00420* (2018).

[25] CHAUDHRY, A., ROHRBACH, M., ELHOSEINY, M., AJANTHAN, T., DOKANIA, P. K., TORR, P. H., AND RANZATO, M. Continual learning with tiny episodic memories.

[26] CHEN, G., QIAO, L., SHI, Y., PENG, P., LI, J., HUANG, T., PU, S., AND TIAN, Y. Learning open set network with discriminative reciprocal points. In *European Conference on Computer Vision* (2020), Springer, pp. 507–522.

[27] CHEN, Q., GE, T., XU, Y., ZHANG, Z., YANG, X., AND GAI, K. Semantic human matting. In *Proceedings of the 26th ACM international conference on Multimedia* (2018), pp. 618–626.

[28] CHEN, Q., LI, D., AND TANG, C.-K. Knn matting. *IEEE transactions on pattern analysis and machine intelligence 35*, 9 (2013), 2175–2188.

[29] CHO, D., TAI, Y.-W., AND KWEON, I. Natural image matting using deep convolutional neural networks. In *European Conference on Computer Vision* (2016), Springer, pp. 626–643.

[30] CHOPRA, S., HADSELL, R., AND LECUN, Y. Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (2005), vol. 1, IEEE, pp. 539–546.

[31] CHUANG, Y.-Y., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. A bayesian approach to digital matting. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* (2001), vol. 2, IEEE, pp. II–II.

[32] CORTES, C., AND VAPNIK, V. Support-vector networks. *Machine learning 20* (1995), 273–297.

[33] CRAMMER, K., DEKEL, O., KESHET, J., SHALEV-SHWARTZ, S., AND SINGER, Y. Online passive aggressive algorithms.

[34] CREUSOT, C., AND MUNAWAR, A. Real-time small obstacle detection on highways using compressive rbm road reconstruction. In *2015 IEEE Intelligent Vehicles Symposium (IV)* (2015), IEEE, pp. 162–167.

[35] DAI, Y., LU, H., AND SHEN, C. Learning affinity-aware upsampling for deep image matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 6841–6850.

[36] DE LANGE, M., ALJUNDI, R., MASANA, M., PARISOT, S., JIA, X., LEONARDIS, A., SLABAUGH, G., AND TUYTELAARS, T. A continual learning survey: Defying forgetting in classification tasks. *IEEE transactions on pattern analysis and machine intelligence 44*, 7 (2021), 3366–3385.

[37] DE LANGE, M., AND TUYTELAARS, T. Continual prototype evolution: Learning online from non-stationary data streams. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 8250–8259.

[38] DE ROSA, R., MENSINK, T., AND CAPUTO, B. Online open world recognition. *arXiv preprint arXiv:1604.02275* (2016).

[39] DENG, J., GUO, J., XUE, N., AND ZAFEIRIOU, S. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019), pp. 4690–4699.

[40] DENOUDEN, T., SALAY, R., CZARNECKI, K., ABDELZAD, V., PHAN, B., AND VERNEKAR, S. Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance. *arXiv preprint arXiv:1812.02765* (2018).

[41] DHAMIJA, A. R., GÜNTHER, M., AND BOULT, T. Reducing network agnostophobia. *Advances in Neural Information Processing Systems 31* (2018).

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

[43] Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. The pascal visual object classes challenge: A retrospective. *International journal of computer vision 111*, 1 (2015), 98–136.

[44] Fei, G., and Liu, B. Breaking the closed world assumption in text classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2016), pp. 506–514.

[45] Fei-Fei, L., Fergus, R., and Perona, P. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence 28*, 4 (2006), 594–611.

[46] Feng, Q., Kang, G., Fan, H., and Yang, Y. Attract or distract: Exploit the margin of open set. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), pp. 7990–7999.

[47] Fernando, C., Banarse, D., Blundell, C., Zwols, Y., Ha, D., Rusu, A. A., Pritzel, A., and Wierstra, D. Pathnet: Evolution channels gradient descent in super neural networks. *arXiv preprint arXiv:1701.08734* (2017).

[48] Fink, M., Shalev-Shwartz, S., Singer, Y., and Ullman, S. Online multiclass learning by interclass hypothesis sharing. In *Proceedings of the 23rd international conference on Machine learning* (2006), pp. 313–320.

[49] Gal, Y., and Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (2016), PMLR, pp. 1050–1059.

[50] GALAMBOS, J. Extreme value theory for applications. In *Extreme Value Theory and Applications: Proceedings of the Conference on Extreme Value Theory and Applications, Volume 1 Gaithersburg Maryland 1993* (1994), Springer, pp. 1–14.

[51] GARG, A., AND ROTH, D. Margin distribution and learning. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)* (2003), pp. 210–217.

[52] GASTAL, E. S., AND OLIVEIRA, M. M. Shared sampling for real-time alpha matting. In *Computer Graphics Forum* (2010), vol. 29, Wiley Online Library, pp. 575–584.

[53] GE, Z., DEMYANOV, S., CHEN, Z., AND GARNAVI, R. Generative openmax for multi-class open set classification. In *British Machine Vision Conference* (2017), BMVA Press.

[54] GENG, C., HUANG, S.-J., AND CHEN, S. Recent advances in open set recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence 43*, 10 (2020), 3614–3631.

[55] GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A., AND BENGIO, Y. Generative adversarial networks. *Communications of the ACM 63*, 11 (2020), 139–144.

[56] GOODFELLOW, I. J., SHLENS, J., AND SZEGEDY, C. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572* (2014).

[57] GRADY, L., SCHIWIETZ, T., AHARON, S., AND WESTERMANN, R. Random walks for interactive alpha-matting. In *Proceedings of VIIP* (2005), vol. 2005, pp. 423–429.

[58] GUNTHER, M., CRUZ, S., RUDD, E. M., AND BOULT, T. E. Toward open-set face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2017), pp. 71–80.

[59] GUO, C., PLEISS, G., SUN, Y., AND WEINBERGER, K. Q. On calibration of modern neural networks. In *International Conference on Machine Learning* (2017), PMLR, pp. 1321–1330.

[60] HASSEN, M., AND CHAN, P. K. Learning a neural-network-based representation for open set recognition. In *Proceedings of the 2020 SIAM International Conference on Data Mining* (2020), SIAM, pp. 154–162.

[61] HE, K., RHEMANN, C., ROTHER, C., TANG, X., AND SUN, J. A global sampling method for alpha matting. In *CVPR 2011* (2011), IEEE, pp. 2049–2056.

[62] HE, K., ZHANG, X., REN, S., AND SUN, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.

[63] HENDRYCKS, D., BASART, S., MAZEIKA, M., ZOU, A., KWON, J., MOSTAJABI, M., STEINHARDT, J., AND SONG, D. Scaling out-of-distribution detection for real-world settings. *ICML* (2022).

[64] HENDRYCKS, D., AND GIMPEL, K. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *Proceedings of International Conference on Learning Representations* (2017).

[65] HENRYDOSS, J., CRUZ, S., RUDD, E. M., GUNTHER, M., AND BOULT, T. E. Incremental open set intrusion recognition using extreme value machine. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (2017), IEEE, pp. 1089–1093.

[66] IOFFE, S., AND SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (2015), PMLR, pp. 448–456.

[67] ISELE, D., AND COSGUN, A. Selective experience replay for lifelong learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (2018), vol. 32.

[68] JAIN, L. P., SCHEIRER, W. J., AND BOULT, T. E. Multi-class open set recognition using probability of inclusion. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III 13* (2014), Springer, pp. 393–409.

[69] JANG, J., NA, B., SHIN, D. H., JI, M., SONG, K., AND MOON, I.-C. Unknown-aware domain adversarial learning for open-set domain adaptation. *Advances in Neural Information Processing Systems 35* (2022), 16755–16767.

[70] JIN, M., KIM, B.-K., AND SONG, W.-J. Adaptive propagation-based color-sampling for alpha matting. *IEEE Transactions on Circuits and Systems for Video Technology 24*, 7 (2014), 1101–1110.

[71] JING, M., LI, J., ZHU, L., DING, Z., LU, K., AND YANG, Y. Balanced open set domain adaptation via centroid alignment. In *Proceedings of the AAAI conference on artificial intelligence* (2021), vol. 35, pp. 8013–8020.

[72] JO, I., KIM, J., KANG, H., KIM, Y.-D., AND CHOI, S. Open set recognition by regularising classifier with fake data generated by generative adversarial networks. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (2018), IEEE, pp. 2686–2690.

[73] JUNG, H., JU, J., JUNG, M., AND KIM, J. Less-forgetting learning in deep neural networks. *arXiv preprint arXiv:1607.00122* (2016).

[74] KARDAN, N., AND STANLEY, K. O. Mitigating fooling with competitive overcomplete output layer neural networks. In *2017 International Joint Conference on Neural Networks (IJCNN)* (2017), IEEE, pp. 518–525.

[75] KE, Z., SUN, J., LI, K., YAN, Q., AND LAU, R. W. Modnet: Real-time trimap-free portrait matting via objective decomposition. In *Proceedings of the AAAI Conference on Artificial Intelligence* (2022), vol. 36, pp. 1140–1147.

[76] KIRKPATRICK, J., PASCANU, R., RABINOWITZ, N., VENESS, J., DESJARDINS, G., RUSU, A. A., MILAN, K., QUAN, J., RAMALHO, T., GRABSKA-BARWINSKA, A., ET AL. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences 114*, 13 (2017), 3521–3526.

[77] KOHONEN, T. Learning vector quantization. In *Self-organizing maps*. Springer, 1995, pp. 175–189.

[78] KUZBORSKIJ, I., ORABONA, F., AND CAPUTO, B. From n to n+ 1: Multiclass transfer incremental learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2013), pp. 3358–3365.

[79] LAKSHMINARAYANAN, B., PRITZEL, A., AND BLUNDELL, C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems 30* (2017).

[80] LEE, P., AND WU, Y. Nonlocal matting. In *CVPR 2011* (2011), IEEE, pp. 2193–2200.

[81] LEVIN, A., LISCHINSKI, D., AND WEISS, Y. A closed-form solution to natural image matting. *IEEE transactions on pattern analysis and machine intelligence 30*, 2 (2007), 228–242.

[82] LEVIN, A., RAV-ACHA, A., AND LISCHINSKI, D. Spectral matting. *IEEE transactions on pattern analysis and machine intelligence 30*, 10 (2008), 1699–1712.

[83] LI, J., MA, S., ZHANG, J., AND TAO, D. Privacy-preserving portrait matting. In *Proceedings of the 29th ACM International Conference on Multimedia* (2021), pp. 3501–3509.

[84] LI, J., ZHANG, J., MAYBANK, S. J., AND TAO, D. Bridging composite and real: towards end-to-end deep image matting. *International Journal of Computer Vision 130*, 2 (2022), 246–266.

[85] LI, J., ZHANG, J., AND TAO, D. Deep automatic natural image matting. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21* (8 2021), Z.-H. Zhou, Ed., International Joint Conferences on Artificial Intelligence Organization, pp. 800–806. Main Track.

[86] LI, Y., AND LU, H. Natural image matting via guided contextual attention. In *Proceedings of the AAAI Conference on Artificial Intelligence* (2020), vol. 34, pp. 11450–11457.

[87] LI, Y., WANG, N., SHI, J., LIU, J., AND HOU, X. Revisiting batch normalization for practical domain adaptation. *arXiv preprint arXiv:1603.04779* (2016).

[88] LI, Z., AND HOIEM, D. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence 40*, 12 (2017), 2935–2947.

[89] LIN, S., RYABTSEV, A., SENGUPTA, S., CURLESS, B. L., SEITZ, S. M., AND KEMELMACHER-SHLIZERMAN, I. Real-time high-resolution background matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 8762–8771.

[90] LIN, T.-Y., MAIRE, M., BELONGIE, S., HAYS, J., PERONA, P., RAMANAN, D., DOLLÁR, P., AND ZITNICK, C. L. Microsoft coco: Common objects in context. In *European conference on computer vision* (2014), Springer, pp. 740–755.

[91] LIU, H., CAO, Z., LONG, M., WANG, J., AND YANG, Q. Separate to adapt: Open set domain adaptation via progressive separation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019), pp. 2927–2936.

[92] LIU, J., GUO, X., AND YUAN, Y. Unknown-oriented learning for open set domain adaptation. In *European Conference on Computer Vision* (2022), Springer, pp. 334–350.

[93] LIU, J., YAO, Y., HOU, W., CUI, M., XIE, X., ZHANG, C., AND HUA, X.-S. Boosting semantic human matting with coarse annotations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 8563–8572.

[94] LIU, Q., XIE, H., ZHANG, S., ZHONG, B., AND JI, R. Long-range feature propagating for natural image matting. In *Proceedings of the 29th ACM International Conference on Multimedia* (2021), pp. 526–534.

[95] LIU, W., LUO, W., LIAN, D., AND GAO, S. Future frame prediction for anomaly detection–a new baseline. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 6536–6545.

[96] LIU, W., WANG, X., OWENS, J., AND LI, Y. Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems 33* (2020), 21464–21475.

[97] LIU, W., WEN, Y., YU, Z., AND YANG, M. Large-margin softmax loss for convolutional neural networks. In *ICML* (2016), vol. 2, p. 7.

[98] LIU, W., ZHANG, Y.-M., LI, X., YU, Z., DAI, B., ZHAO, T., AND SONG, L. Deep hyperspherical learning. *Advances in neural information processing systems 30* (2017).

[99] LIU, Y., LI, H., AND WANG, X. Rethinking feature discrimination and polymerization for large-scale recognition. *arXiv preprint arXiv:1710.00870* (2017).

[100] LIU, Z., CHEN, G., LI, Z., KANG, Y., QU, S., AND JIANG, C. Psdc: A prototype-based shared-dummy classifier model for open-set domain adaptation. *IEEE Transactions on Cybernetics* (2022).

[101] LOPEZ-PAZ, D., AND RANZATO, M. Gradient episodic memory for continual learning. *Advances in neural information processing systems 30* (2017).

[102] LU, H., DAI, Y., SHEN, C., AND XU, S. Indices matter: Learning to index for deep image matting. In *Proceedings of the IEEE International Conference on Computer Vision* (2019), pp. 3266–3275.

[103] LUO, Y., WANG, Z., HUANG, Z., AND BAKTASHMOTLAGH, M. Progressive graph learning for open-set domain adaptation. In *International Conference on Machine Learning* (2020), PMLR, pp. 6468–6478.

[104] LUTZ, S., AMPLIANITIS, K., AND SMOLIC, A. Alphagan: Generative adversarial networks for natural image matting.

[105] MADDOX, W. J., IZMAILOV, P., GARIPOV, T., VETROV, D. P., AND WILSON, A. G. A simple baseline for bayesian uncertainty in deep learning. *Advances in Neural Information Processing Systems 32* (2019).

[106] MALLYA, A., AND LAZEBNIK, S. Packnet: Adding multiple tasks to a single network by iterative pruning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (2018), pp. 7765–7773.

[107] MENDES JÚNIOR, P. R., DE SOUZA, R. M., WERNECK, R. D. O., STEIN, B. V., PAZINATO, D. V., DE ALMEIDA, W. R., PENATTI, O. A., TORRES, R. D. S., AND ROCHA, A. Nearest neighbors distance ratio open-set classifier. *Machine Learning 106*, 3 (2017), 359–386.

[108] MENSINK, T., VERBEEK, J., PERRONNIN, F., AND CSURKA, G. Distance-based image classification: Generalizing to new classes at near-zero cost. *IEEE transactions on pattern analysis and machine intelligence 35*, 11 (2013), 2624–2637.

[109] MILLER, D., SUNDERHAUF, N., MILFORD, M., AND DAYOUB, F. Class anchor clustering: A loss for distance-based open set recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2021), pp. 3570–3578.

[110] MONTAVON, G., ORR, G., AND MÜLLER, K.-R. *Neural networks: tricks of the trade*, vol. 7700. springer, 2012.

[111] MUHLBAIER, M. D., TOPALIS, A., AND POLIKAR, R. Learn$^{\{++\}}$. nc: Combining ensemble of classifiers with dynamically weighted consult-and-vote for efficient incremental learning of new classes. *IEEE transactions on neural networks 20*, 1 (2008), 152–168.

[112] NEAL, L., OLSON, M., FERN, X., WONG, W.-K., AND LI, F. Open set learning with counterfactual images. In *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 613–628.

[113] NEIRA, M. A. C., JÚNIOR, P. R. M., ROCHA, A., AND TORRES, R. D. S. Data-fusion techniques for open-set recognition problems. *IEEE Access 6* (2018), 21242–21265.

[114] NETZER, Y., WANG, T., COATES, A., BISSACCO, A., WU, B., AND NG, A. Y. Reading digits in natural images with unsupervised feature learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011* (2011).

[115] NGUYEN, A., YOSINSKI, J., AND CLUNE, J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 427–436.

[116] NGUYEN, C. V., LI, Y., BUI, T. D., AND TURNER, R. E. Variational continual learning. In *International Conference on Learning Representations* (2018).

[117] ORESHKIN, B., RODRÍGUEZ LÓPEZ, P., AND LACOSTE, A. Tadam: Task dependent adaptive metric for improved few-shot learning. *Advances in neural information processing systems 31* (2018).

[118] OZA, P., AND PATEL, V. M. C2ae: Class conditioned auto-encoder for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 2307–2316.

[119] PAN, Y., YAO, T., LI, Y., NGO, C.-W., AND MEI, T. Exploring category-agnostic clusters for open-set domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020), pp. 13867–13875.

[120] PANAREDA BUSTO, P., AND GALL, J. Open set domain adaptation. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 754–763.

[121] PARK, G., SON, S., YOO, J., KIM, S., AND KWAK, N. Matteformer: Transformer-based image matting via prior-tokens. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 11696–11706.

[122] PELCKMANS, K., SUYKENS, J., AND MOOR, B. A risk minimization principle for a class of parzen estimators. *Advances in Neural Information Processing Systems 20* (2007).

[123] PENG, J., LI, L., AND TANG, Y. Y. Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images. *IEEE transactions on neural networks and learning systems 30*, 6 (2018), 1790–1802.

[124] QIAN, R. J., AND SEZAN, M. I. Video background replacement without a blue screen. In *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)* (1999), vol. 4, pp. 143–146 vol.4.

[125] QIAO, Y., LIU, Y., YANG, X., ZHOU, D., XU, M., ZHANG, Q., AND WEI, X. Attention-guided hierarchical structure aggregation for image matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 13676–13685.

[126] RANJAN, R., CASTILLO, C. D., AND CHELLAPPA, R. L2-constrained softmax loss for discriminative face verification. *arXiv preprint arXiv:1703.09507* (2017).

[127] REBUFFI, S.-A., KOLESNIKOV, A., SPERL, G., AND LAMPERT, C. H. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (2017), pp. 2001–2010.

[128] REYZIN, L., AND SCHAPIRE, R. E. How boosting the margin can also boost classifier complexity. In *Proceedings of the 23rd international conference on Machine learning* (2006), pp. 753–760.

[129] RISTIN, M., GUILLAUMIN, M., GALL, J., AND VAN GOOL, L. Incremental learning of ncm forests for large-scale image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2014), pp. 3654–3661.

[130] ROLNICK, D., AHUJA, A., SCHWARZ, J., LILLICRAP, T., AND WAYNE, G. Experience replay for continual learning. *Advances in Neural Information Processing Systems 32* (2019).

[131] ROZSA, A., GÜNTHER, M., AND BOULT, T. E. Adversarial robustness: Softmax versus openmax. *arXiv preprint arXiv:1708.01697* (2017).

[132] RUBINSTEIN, R., BRUCKSTEIN, A. M., AND ELAD, M. Dictionaries for sparse representation modeling. *Proceedings of the IEEE 98*, 6 (2010), 1045–1057.

[133] RUDD, E. M., JAIN, L. P., SCHEIRER, W. J., AND BOULT, T. E. The extreme value machine. *IEEE transactions on pattern analysis and machine intelligence 40*, 3 (2017), 762–768.

[134] RUSU, A. A., RABINOWITZ, N. C., DESJARDINS, G., SOYER, H., KIRKPATRICK, J., KAVUKCUOGLU, K., PASCANU, R., AND HADSELL, R. Progressive neural networks. *arXiv preprint arXiv:1606.04671* (2016).

[135] SAITO, K., AND SAENKO, K. Ovanet: One-vs-all network for universal domain adaptation. In *Proceedings of the ieee/cvf international conference on computer vision* (2021), pp. 9000–9009.

[136] SAITO, K., YAMAMOTO, S., USHIKU, Y., AND HARADA, T. Open set domain adaptation by backpropagation. In *Proceedings of the European conference on computer vision (ECCV)* (2018), pp. 153–168.

[137] SCHEIRER, W. J., DE REZENDE ROCHA, A., SAPKOTA, A., AND BOULT, T. E. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence 35*, 7 (2012), 1757–1772.

[138] SCHEIRER, W. J., JAIN, L. P., AND BOULT, T. E. Probability models for open set recognition. *IEEE transactions on pattern analysis and machine intelligence 36*, 11 (2014), 2317–2324.

[139] SCHERREIK, M. D., AND RIGLING, B. D. Open set recognition for automatic target classification with rejection. *IEEE Transactions on Aerospace and Electronic Systems 52*, 2 (2016), 632–642.

[140] SCHROFF, F., KALENICHENKO, D., AND PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 815–823.

[141] SCHULTZ, M., AND JOACHIMS, T. Learning a distance metric from relative comparisons. *Advances in neural information processing systems 16* (2003).

[142] SEIDENSCHWARZ, J. D., ELEZI, I., AND LEAL-TAIXÉ, L. Learning intra-batch connections for deep metric learning. In *International Conference on Machine Learning* (2021), PMLR, pp. 9410–9421.

[143] SENGUPTA, S., JAYARAM, V., CURLESS, B., SEITZ, S. M., AND KEMELMACHER-SHLIZERMAN, I. Background matting: The world is your green screen. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 2291–2300.

[144] Serra, J., Suris, D., Miron, M., and Karatzoglou, A. Overcoming catastrophic forgetting with hard attention to the task. In *International Conference on Machine Learning* (2018), PMLR, pp. 4548–4557.

[145] Shaban, A., Bansal, S., Liu, Z., Essa, I., and Boots, B. One-shot learning for semantic segmentation. In *BMVC* (2017).

[146] Shahrian, E., Rajan, D., Price, B., and Cohen, S. Improving image matting using comprehensive sampling sets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2013), pp. 636–643.

[147] Shin, H., Lee, J. K., Kim, J., and Kim, J. Continual learning with deep generative replay. *Advances in neural information processing systems 30* (2017).

[148] Shu, L., Xu, H., and Liu, B. Doc: Deep open classification of text documents. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (2017), pp. 2911–2916.

[149] Shu, L., Xu, H., and Liu, B. Unseen class discovery in open-world classification. *arXiv preprint arXiv:1801.05609* (2018).

[150] Shu, Y., Shi, Y., Wang, Y., Zou, Y., Yuan, Q., and Tian, Y. Odn: Opening the deep network for open-set action recognition. In *2018 IEEE international conference on multimedia and expo (ICME)* (2018), IEEE, pp. 1–6.

[151] Silver, D. L., and Mercer, R. E. The task rehearsal method of life-long learning: Overcoming impoverished data. In *Conference of the Canadian Society for Computational Studies of Intelligence* (2002), Springer, pp. 90–101.

[152] Snell, J., Swersky, K., and Zemel, R. Prototypical networks for few-shot learning. *Advances in neural information processing systems 30* (2017).

[153] Sun, J., Jia, J., Tang, C.-K., and Shum, H.-Y. Poisson matting. In *ACM SIGGRAPH 2004 Papers*. 2004, pp. 315–321.

[154] SUN, J., KE, Z., XU, K., SHAO, F., ZHANG, L., LU, H., AND LAU, R. W. Semantics-adding flaw-erasing network for semantic human matting. In *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022* (2022), BMVA Press.

[155] SUN, Y., CHEN, Y., WANG, X., AND TANG, X. Deep learning face representation by joint identification-verification. *Advances in neural information processing systems 27* (2014).

[156] SUN, Y., TANG, C.-K., AND TAI, Y.-W. Semantic image matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 11120–11129.

[157] TANG, J., AKSOY, Y., OZTIRELI, C., GROSS, M., AND AYDIN, T. O. Learning-based sampling for natural image matting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), pp. 3055–3063.

[158] THULASIDASAN, S., CHENNUPATI, G., BILMES, J. A., BHATTACHARYA, T., AND MICHALAK, S. On mixup training: Improved calibration and predictive uncertainty for deep neural networks. *Advances in Neural Information Processing Systems 32* (2019).

[159] VARETO, R., SILVA, S., COSTA, F., AND SCHWARTZ, W. R. Towards open-set face recognition using hashing functions. In *2017 IEEE international joint conference on biometrics (IJCB)* (2017), IEEE, pp. 634–641.

[160] VENKATARAM, V. M. *Open set text classification using neural networks*. University of Colorado Colorado Springs, 2018.

[161] VIGNOTTO, E., AND ENGELKE, S. Extreme value theory for open set classification–gpd and gev classifiers. *arXiv preprint arXiv:1808.09902* (2018).

[162] VINYALS, O., BLUNDELL, C., LILLICRAP, T., WIERSTRA, D., ET AL. Matching networks for one shot learning. *Advances in neural information processing systems 29* (2016).

[163] WANG, F., CHENG, J., LIU, W., AND LIU, H. Additive margin softmax for face verification. *IEEE Signal Processing Letters 25*, 7 (2018), 926–930.

[164] WANG, F., XIANG, X., CHENG, J., AND YUILLE, A. L. Normface: L2 hypersphere embedding for face verification. In *Proceedings of the 25th ACM international conference on Multimedia* (2017), pp. 1041–1049.

[165] WANG, H., WANG, Y., ZHOU, Z., JI, X., GONG, D., ZHOU, J., LI, Z., AND LIU, W. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 5265–5274.

[166] WANG, J., AND COHEN, M. F. An iterative optimization approach for unified image segmentation and matting. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* (2005), vol. 2, IEEE, pp. 936–943.

[167] WANG, J., AND COHEN, M. F. Optimized color sampling for robust matting. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (2007), IEEE, pp. 1–8.

[168] WANG, L., LU, H., WANG, Y., FENG, M., WANG, D., YIN, B., AND RUAN, X. Learning to detect salient objects with image-level supervision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 136–145.

[169] WANG, Y.-X., GIRSHICK, R., HEBERT, M., AND HARIHARAN, B. Low-shot learning from imaginary data. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 7278–7286.

[170] WEI, T., CHEN, D., ZHOU, W., LIAO, J., ZHAO, H., ZHANG, W., AND YU, N. Improved image matting via real-time user clicks and uncertainty estimation. In

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2021), pp. 15374–15383.

[171] WEN, Y., ZHANG, K., LI, Z., AND QIAO, Y. A discriminative feature learning approach for deep face recognition. In *European conference on computer vision* (2016), Springer, pp. 499–515.

[172] WIESLER, S., AND NEY, H. A convergence analysis of log-linear training. *Advances in Neural Information Processing Systems 24* (2011).

[173] WRIGHT, J., MA, Y., MAIRAL, J., SAPIRO, G., HUANG, T. S., AND YAN, S. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE 98*, 6 (2010), 1031–1044.

[174] WRIGHT, J., YANG, A. Y., GANESH, A., SASTRY, S. S., AND MA, Y. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence 31*, 2 (2008), 210–227.

[175] XIAO, C., LIU, M., XIAO, D., DONG, Z., AND MA, K.-L. Fast closed-form matting using a hierarchical data structure. *IEEE Transactions on Circuits and Systems for Video Technology 24*, 1 (2013), 49–62.

[176] XU, J., AND ZHU, Z. Reinforced continual learning. *Advances in Neural Information Processing Systems 31* (2018).

[177] XU, N., PRICE, B., COHEN, S., AND HUANG, T. Deep image matting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 2970–2979.

[178] XU, R., LIU, P., ZHANG, Y., CAI, F., WANG, J., LIANG, S., YING, H., AND YIN, J. Joint partial optimal transport for open set domain adaptation. In *IJCAI* (2020), pp. 2540–2546.

[179] XU, Y., LIU, B., QUAN, Y., AND JI, H. Unsupervised deep background matting using deep matte prior. *IEEE Transactions on Circuits and Systems for Video Technology* (2021).

[180] YANG, H.-M., ZHANG, X.-Y., YIN, F., AND LIU, C.-L. Robust classification with convolutional prototype learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 3474–3482.

[181] YANG, H.-M., ZHANG, X.-Y., YIN, F., YANG, Q., AND LIU, C.-L. Convolutional prototype network for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[182] YANG, Y., GAO, R., AND XU, Q. Out-of-distribution detection with semantic mismatch under masking. In *European Conference on Computer Vision* (2022), Springer, pp. 373–390.

[183] YANG, Y., HOU, C., LANG, Y., GUAN, D., HUANG, D., AND XU, J. Open-set human activity recognition based on micro-doppler signatures. *Pattern Recognition 85* (2019), 60–69.

[184] YEH, T., AND DARRELL, T. Dynamic visual category learning. In *2008 IEEE Conference on Computer Vision and Pattern Recognition* (2008), IEEE, pp. 1–8.

[185] YI, D., LEI, Z., LIAO, S., AND LI, S. Z. Deep metric learning for person re-identification. In *2014 22nd international conference on pattern recognition* (2014), IEEE, pp. 34–39.

[186] YOSHIHASHI, R., SHAO, W., KAWAKAMI, R., YOU, S., IIDA, M., AND NAEMURA, T. Classification-reconstruction learning for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 4016–4025.

[187] YU, B., AND TAO, D. Deep metric learning with tuplet margin loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), pp. 6490–6499.

[188] YU, F., ZHANG, Y., SONG, S., SEFF, A., AND XIAO, J. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop.

[189] YU, H., XU, N., HUANG, Z., ZHOU, Y., AND SHI, H. High-resolution deep image matting. In *Proceedings of the AAAI Conference on Artificial Intelligence* (2021), vol. 35, pp. 3217–3224.

[190] YU, Q., IRIE, G., AND AIZAWA, K. Self-labeling framework for open-set domain adaptation with few labeled samples. *IEEE Transactions on Multimedia* (2023).

[191] YU, Q., ZHANG, J., ZHANG, H., WANG, Y., LIN, Z., XU, N., BAI, Y., AND YUILLE, A. Mask guided matting via progressive refinement network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 1154–1163.

[192] YU, Y., QU, W.-Y., LI, N., AND GUO, Z. Open-category classification by adversarial sample generation. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence* (2017), pp. 3357–3363.

[193] ZAEEMZADEH, A., BISAGNO, N., SAMBUGARO, Z., CONCI, N., RAHNAVARD, N., AND SHAH, M. Out-of-distribution detection using union of 1-dimensional subspaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 9452–9461.

[194] ZENATI, H., FOO, C. S., LECOUAT, B., MANEK, G., AND CHANDRASEKHAR, V. R. Efficient gan-based anomaly detection. *arXiv preprint arXiv:1802.06222* (2018).

[195] ZENKE, F., POOLE, B., AND GANGULI, S. Continual learning through synaptic intelligence. In *International Conference on Machine Learning* (2017), PMLR, pp. 3987–3995.

[196] ZENO, C., GOLAN, I., HOFFER, E., AND SOUDRY, D. Task-agnostic continual learning using online variational bayes with fixed-point updates. *Neural Computation 33*, 11 (2021), 3139–3177.

[197] ZHANG, H., CISSE, M., DAUPHIN, Y. N., AND LOPEZ-PAZ, D. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412* (2017).

[198] ZHANG, H., AND PATEL, V. M. Sparse representation-based open set recognition. *IEEE transactions on pattern analysis and machine intelligence 39*, 8 (2016), 1690–1696.

[199] ZHANG, J., ZHANG, J., GHOSH, S., LI, D., TASCI, S., HECK, L., ZHANG, H., AND KUO, C.-C. J. Class-incremental learning via deep model consolidation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2020), pp. 1131–1140.

[200] ZHANG, Y., GONG, L., FAN, L., REN, P., HUANG, Q., BAO, H., AND XU, W. A late fusion cnn for digital matting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), pp. 7469–7478.

[201] ZHENG, Y., PAL, D. K., AND SAVVIDES, M. Ring loss: Convex feature normalization for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 5089–5097.

[202] ZHOU, Y. Rethinking reconstruction autoencoder-based out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 7379–7387.

[203] ZHOU, Y., LARADJI, I. H., ZHOU, L., AND NOWROUZEZAHRAI, D. Osm: An open set matting framework with ood detection and few-shot learning. In *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022* (2022), BMVA Press.