Task-Adaptive Negative Envision for Few-Shot Open-Set Recognition
Supplementary Materials

1. Baseline Implementations

We implement some of the baseline methods, which are marked with * in the main paper. For fair comparison, all the re-implemented methods use the same feature extractor (ResNet12) as ours, which is first pre-trained on the base set.

Dynamic: Our ATT-G and SEMAN-G negative generators are built upon Dynamic [10]. Dynamic [10] calibrates few-shot prototypes using base class classifier. We re-implemented Dynamic using the standard Transformer attention block. We compare this baseline in Tab. 3 in the main paper for GFSOR performance (i.e., DynamicFSL) , where we calculate its accuracy in the normal way, while for AUROC, we take $\mathrm{argmax}_{y\in\ \mathcal{C}^f\cup\mathcal{C}^*} p(x|y)$ as the rejection score. We also evaluate its performance on CIFAR-FS and FC-100 in Tab. 6, where we take $\mathrm{argmax}_{y\in\ \mathcal{C}^f} p(x|y)$ as the rejection score.

PEELER: Our training and sampling strategy are built upon PEELER [24], and hence pick it as one threshold-based method for comparison in Fig. 3 and Tab. 1. In detail, we implement PEELER using our pre-trained feature extractor, and then perform meta-training with PEELER's open-set loss. Then rejection score is calculated as $\mathrm{argmax}_{y\in\ \mathcal{C}^f} p(x|y)$.

Dynamic+PEELER. To compare to ATT-G more fairly, we combine PEELER with Dynamic as another threshold-based approach in Fig. 3 and Tab. 1. Specifically, we use PEELER's loss and training strategy on top of the calibrated few-shot prototypes. Then the rejection score is calculated as $\mathrm{argmax}_{y\in\ \mathcal{C}^f} p(x|y)$.

CounterFactual. CounterFactual [27] is a generative open-set recognition method. To apply in our FSOR setting, we first train the GAN network on base set and use the support set $\mathcal{S}$ to synthesize fake images. The averaged fake image feature is used as the negative prototype for open-set recognition.

OpenMax [1]. For experiments on CIFAR-FS and FC100, we follow [20] to fit Weibull models over training tasks using the predicted class scores (i.e., logits). Then the mean activation vectors are used for negative detection.

2. Dataset Descriptions

For the brevity of description, we follow the order $\{$ base, novel (validation), novel (testing) $\}$ to describe the class splits. The base class split is for training the few-shot model. We use the novel (validation) class split for model selection, and report the evaluation results on the novel (testing) class split.

MiniImageNet [43] and TieredImageNet [33] are two subsets of ImageNet-1k [4]. MiniImageNet contains 100 object classes and each class contains 600 images. The 100 classes are split into (64, 16, 20). TieredImageNet contains the full set of 608 object classes. The 608 classes are split into (351, 97, 160). All the images are resized to 84$\times$84. TieredImageNet offers more challenging class splits than MiniImageNet that are based on object class hierarchy. In our experiment, we use both datasets to evaluate FSOR tasks, and use MiniImageNet to additionally evaluate GFSOR tasks. Specifically, MiniImageNet offers additional 300 base-class images for testing. For GFSOR evaluation, we use this base-class test split to sample base-class positive queries.

CIFAR-FS [2] and **FC100** [29] are two different splits of CIFAR100 [**?**] with 100 object classes and each class has 600 images. The images are at a resolution of 32 $\times$ 32. CIFAR-FS [2] has the class splits (64,16,20), while FC100 [29] has the class splits (60,20,20). In particular, FC100 offers more challenging class splits that are based on object class hierarchy. We use both datasets to further evaluate FSOR tasks.

3. Figure Details

For Fig. 1, we sample two tasks from MiniImageNet. Within each task, we sample three few-shot classes and select one training sample to estimate class prototype for each class. We then select another three classes whose samples are used as negative queries in both tasks. In Fig. 1(up), for each class, we calculate the cosine similarity score as detection score between the class prototype and all of its positive queries. The mean and standard deviation of detection scores are represented by the height of bar and half length of black line respectively. Similarly, we calculate the cosine

Table S1. Summary of Symbol Notation

Variable	Definition	Note
$\mathcal{T}$	a few-shot open-set recognition task	
$\mathcal{T}^*$	a generalized few-shot open-set recognition task	
$\mathcal{C}^f$	the set of few-shot classes	
$\mathcal{C}^*$	the set of many-shot classes	
$\mathcal{C}^n$	the set of negative classes (simulate unknown sources)	
$\mathcal{C}^B/\mathcal{C}^N$	base/novel class set	class split in experiment
$\mathcal{D}^B/\mathcal{D}^N$	base/novel class dataset	dataset split in experiment
$\mathcal{S}$	training samples (support) set of few-shot classes	
$\mathcal{S}_c$	training samples (support) set of few-shot class c	(character sub-index)
$\mathcal{Q}^f/\mathcal{Q}^*/\mathcal{Q}^n$	testing samples (query) set of few-shot/many-shot/negative classes	
$\mathbf{P}^f$	few-shot prototype vectors (average the support feature vectors)	
$\mathbf{P}^*$	many-shot prototype vectors (classifier weights after pretraining)	
$\mathbf{Z}^f/\mathbf{Z}^*$	semantic-visual vectors for few/many-shot classes	
$\mathbf{p}_c$	a prototype feature of class c	
$\mathbf{p}^-$	a negative prototype feature	
$\mathbf{e}_c$	word embedding (semantic representation) of class c	
$\mathbf{z}_c$	semantic-visual vector for class c	concatenate $\mathbf{p}_c$ and $\mathbf{e}_c$
$\mathbf{K}_q, \mathbf{K}_k, \mathbf{K}_v$	learnable kernels for attention calculation	
$\mathbf{A}_{((x,y))}$	the unnormalized attention weights between x and y	
s	a support sample	
q	a query sample	its label is y_q
f	feature extractor function, extract one feature vector for each image	
f_s	distance function, calculate similarity between two feature vectors	
f_n	an MLP for negative prototype generation	
f_g	gating function used in ATT-G and SEMAN-G	
g_n	general representation for the negative prototype generation function	$g_{n,i}$ the i-th function
θ_m	manual threshold for negative query detection	
θ_a	automatic/task-adaptive threshold	$\theta_{a,i}$ the i-th threshold
τ	a margin between rejection score and positive detection scores	the margin value is positive
$\odot$	element-wise multiplication	
ϕ	element-wise sigmoid operation	
σ	row-wise softmax operation	

Note: the numerical sub-index used in the paper is for the definition of conjugate tasks and their related calculation

similarity between the prototype and all negative queries, and calculate the mean and standard deviation. Within Fig. 1(down), for each subplot, we plot a circle centered at a negative query and the negative query will be classified as a positive class if a few-shot class prototype is in the circle.

Table S2. 5-way FSOR results CIFAR-FS. *: our implementation.

Algorithm	1-shot		5-shot	
	Acc	AUROC	Acc	AUROC
OpenMax [1]*	71.65±0.65	50.21±0.07	85.66±0.48	75.78±0.47
CounterFactural [27]*	71.71±0.65	72.57±0.61	85.71±0.45	80.44±0.37
PEELER [24]*	71.47±0.67	71.28±0.57	85.46±0.47	75.97±0.33
Dynamic [10]*	71.56±0.67	66.89±0.52	85.78±0.49	76.03±0.37
ATT-G (ours)	72.43±0.65	76.72±0.59	86.52±0.49	84.64±0.38
SEMAN-G (ours)	74.55±0.65	78.10±0.58	86.71±0.47	86.47±0.37

Table S3. 5-way FSOR results on FC100. *: our implementation.

Algorithm	1-shot		5-shot	
	Acc	AUROC	Acc	AUROC
OpenMax [1]*	44.70±0.60	50.10±0.11	60.11±0.62	57.78±0.44
CounterFactural [27]*	44.53±0.60	57.20±0.47	61.12±0.60	62.35±0.45
PEELER [24]*	44.45±0.57	55.86±0.44	60.86±0.59	61.07±0.40
Dynamic [10]*	44.88±0.59	55.62±0.54	60.45±0.61	59.01±0.52
ATT-G (ours)	45.11±0.60	59.55±0.57	61.18±0.61	63.34±0.50
SEMAN-G (ours)	46.01±0.60	59.73±0.53	62.18±0.57	64.46±0.50

4. Terminology & Symbol Summary

Throughout the paper explanation, we interchangeably use the terms *unknown* and *negative*, *known* and *positive*. The *unknown* source is simulated by selecting negative classes.

During meta-training, we build a FSOR task by sampling $\mathcal{C}^f$ and $\mathcal{C}^n$ from $\mathcal{C}^B$ where $\mathcal{C}^n \cap \mathcal{C}^f = \varnothing$. To build a GFSOR task, we split the $\mathcal{C}^B$ into $\mathcal{C}^f$, $\mathcal{C}^n$ and $\mathcal{C}^*$, *e.g.*, on MiniImagenet, to sample a 5-way GFSOR task $\mathcal{T}^*$, we have $|\mathcal{C}^*| = 54$ while $|\mathcal{C}^n| = |\mathcal{C}^f| = 5$.

During meta-testing, we sample $\mathcal{C}^f$ and $\mathcal{C}^n$ from novel classes $\mathcal{C}^N$ where $\mathcal{C}^N \cap \mathcal{C}^B = \varnothing$. As such, we use $\mathcal{C}^B$ as $\mathcal{C}^*$ and thus interchangeably use the term base classes and many-shot classes.

We summarize the symbols in Table. S1.

5. More Experiments

We extend Table 5&6 in the main paper and provide the full results on CIFAR-FS and FC100 in Table S2&S3 respectively.

References

[1] Abhijit Bendale and Terrance E. Boult. Towards open set deep networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2016. 1, 3

[2] Luca Bertinetto, Joao F. Henriques, Philip Torr, and Andrea Vedaldi. Meta-learning with differentiable closed-form solvers. In *Int. Conf. Learn. Represent.*, 2019. 1

[3] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013.

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 1

[5] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[6] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2017.

[7] Sebastian Flennerhag, Andrei A Rusu, Razvan Pascanu, Francesco Visin, Hujun Yin, and Raia Hadsell. Meta-learning with warped gradient descent. *arXiv preprint arXiv:1909.00025*, 2019.

[8] ZongYuan Ge, Sergey Demyanov, Zetao Chen, and Rahil Garnavi. Generative openmax for multi-class open set classification. *arXiv preprint arXiv:1707.07418*, 2017.

[9] Spyros Gidaris, Andrei Bursuc, Nikos Komodakis, Patrick Pérez, and Matthieu Cord. Boosting few-shot visual learning with self-supervision. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8059–8068, 2019.

[10] Spyros Gidaris and Nikos Komodakis. Dynamic few-shot visual learning without forgetting. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018. 1, 3

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[12] Guangxing Han, Yicheng He, Shiyuan Huang, Jiawei Ma, and Shih-Fu Chang. Query adaptive few-shot object detection with heterogeneous graph convolutional

networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3263–3272, October 2021.

[13] Guangxing Han, Shiyuan Huang, Jiawei Ma, Yicheng He, and Shih-Fu Chang. Meta faster r-cnn: Towards accurate few-shot object detection with attentive feature alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.

[14] Guangxing Han, Jiawei Ma, Shiyuan Huang, Long Chen, and Shih-Fu Chang. Few-shot object detection with fully cross-transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[15] Guangxing Han, Xuan Zhang, and Chongrong Li. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3360–3364, 2017.

[16] Guangxing Han, Xuan Zhang, and Chongrong Li. Revisiting faster r-cnn: A deeper look at region proposal network. In *International Conference on Neural Information Processing*, pages 14–24, 2017.

[17] Guangxing Han, Xuan Zhang, and Chongrong Li. Semi-supervised dff: Decoupling detection and feature flow for video object detectors. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 1811–1819, 2018.

[18] De-Chuan Zhan Han-Jia Ye, Hexiang Hu and Fei Sha. Learning adaptive classifiers synthesis for generalized few-shot learning. *arXiv preprint arXiv:1906.02944*, 2019.

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

[20] Minki Jeong, Seokeon Choi, and Changick Kim. Few-shot open-set recognition by transformation consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12566–12575, 2021. 1

[21] Kimin Lee, Honglak Lee, Kibok Lee, and Jinwoo Shin. Training confidence-calibrated classifiers for detecting out-of-distribution samples. *arXiv preprint arXiv:1711.09325*, 2017.

[22] Aoxue Li, Weiran Huang, Xu Lan, Jiashi Feng, Zhenguo Li, and Liwei Wang. Boosting few-shot learning with adaptive margin loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12576–12584, 2020.

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

[24] Bo Liu, Hao Kang, Haoxiang Li, Gang Hua, and Nuno Vasconcelos. Few-shot open-set recognition using meta-learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1, 3

[25] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[26] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[27] Lawrence Neal, Matthew Olson, Xiaoli Fern, Weng-Keen Wong, and Fuxin Li. Open set learning with counterfactual images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. 1, 3

[28] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018.

[29] Boris Oreshkin, Pau Rodríguez López, and Alexandre Lacoste. Tadam: Task dependent adaptive metric for improved few-shot learning. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31, pages 721–731. Curran Associates, Inc., 2018. 1

[30] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[31] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[32] Mengye Ren, Renjie Liao, Ethan Fetaya, and Richard S Zemel. Incremental few-shot learning with attention attractor networks. *arXiv preprint arXiv:1810.07218*, 2018.

[33] Mengye Ren, Sachin Ravi, Eleni Triantafillou, Jake Snell, Kevin Swersky, Josh B. Tenenbaum, Hugo Larochelle, and Richard S. Zemel. Meta-learning for semi-supervised few-shot classification. In *iclr*, 2018. 1

[34] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein,

et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[35] Walter J. Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E. Boult. Toward open set recognition. In *IEEE Trans. Pattern Anal. Mach. Intell.*, number 7, pages 1757–1772, 2013.

[36] Patrick Schlachter, Yiwen Liao, and Bin Yang. Open-set recognition using intra-class splitting. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–5. IEEE, 2019.

[37] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Adv. Neural Inform. Process. Syst.*, 2017.

[38] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 403–412, 2019.

[39] Xin Sun, Zhenning Yang, Chi Zhang, Keck-Voon Ling, and Guohao Peng. Conditional gaussian distribution learning for open set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13480–13489, 2020.

[40] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018.

[41] Yonglong Tian, Yue Wang, Dilip Krishnan, Joshua B Tenenbaum, and Phillip Isola. Rethinking few-shot image classification: a good embedding is all you need? *arXiv preprint arXiv:2003.11539*, 2020.

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[43] Oriol Vinyals, Charles Blundell, Tim Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *Adv. Neural Inform. Process. Syst.*, 2016. 1

[44] Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. Learning to model the tail. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 7032–7042, 2017.

[45] Chen Xing, Negar Rostamzadeh, Boris N Oreshkin, and Pedro O Pinheiro. Adaptive cross-modal few-shot learning. *arXiv preprint arXiv:1902.07104*, 2019.

[46] Han-Jia Ye, Hexiang Hu, De-Chuan Zhan, and Fei Sha. Learning classifier synthesis for generalized few-shot learning. 2019.

[47] Han-Jia Ye, Hexiang Hu, De-Chuan Zhan, and Fei Sha. Few-shot learning via embedding adaptation with set-to-set functions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8808–8817, 2020.

[48] Nikolaos-Antonios Ypsilantis, Noa Garcia, Guangxing Han, Sarah Ibrahimi, Nanne Van Noord, and Giorgos Tolias. The met dataset: Instance-level recognition for artworks. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

[49] Da-Wei Zhou, Han-Jia Ye, and De-Chuan Zhan. Learning placeholders for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2021.