Weight Factorization Based Incremental Learning in Generalized Few Shot Segmentation

Anuska Roy and Viswanath Gopalakrishnan
International Institute of Information Technology, Bangalore, India

Keywords: Few Shot Learning, Meta Learning, Semantic Segmentation.

Abstract: Generalized Few-shot Semantic Segmentation (GFSS) targets to segment novel object categories using a few annotated examples after learning the segmentation on a set of base classes. A typical GFSS training involves two stages - base class learning followed by novel class addition and learning. While existing methods have shown promise, they often struggle when novel classes are significant in number. Most current approaches freeze the encoder backbone to retain base class accuracy; however, freezing the encoder backbone can potentially impede the assimilation of novel information from the new classes. To address this challenge, we propose to use an incremental learning strategy in GFSS for learning both encoder backbone and novel class prototypes. Inspired by the recent success of Low Rank Adaptation techniques (LoRA), we introduce incremental learning to the GFSS encoder backbone with a novel weight factorization method. Our newly proposed rank adaptive weight merging strategy is sensitive to the varying degrees of novelty assimilated across various layers of the encoder backbone. In our work, we also introduce the incremental learning strategy to class prototype learning for novel categories. Our extensive experiments on Pascal-5^i and COCO-20^i databases showcase the effectiveness of incremental learning, especially when the novel classes outnumber base classes. With our proposed Weight Factorization based Incremental Learning (WFIL) method, a new set of state-of-the-art accuracy values is established in Generalized Few-shot Semantic Segmentation.

1 INTRODUCTION

Supervised training of semantic segmentation (SS) requires large labelled training data. This is often challenging in many scenarios since the labelling process is quite labour intensive due to the pixel level labelling requirements. Few-shot Semantic Segmentation (FSS) methods address this issue to a certain extent with the design of architectures that utilizes the labels from a limited set of support images to perform segmentation on a query image. Generalized few-shot Semantic Segmentation (GFSS) assumes a more realistic setting where segmentation is initially learnt on a set of base classes with large number of labelled data. GFSS further adapts the network to learn segmentation on a set of novel classes with limited labels with the objective of performing accurate segmentation on both base and novel classes.

There are primarily two challenges involved with GFSS training - i) Achieving good segmentation accuracies on novel classes with limited labels ii) Holding the accuracy of base classes while learning novel classes. Prior works in GFSS (Tian et al., 2022a), (Liu et al., 2023b),(Huang et al., 2023),(Liu et al., 2023a) have focussed on avoiding detoriation of accuracy of base classes by freezing the encoder backbone network while learning novel classes. This limits the capacity of encoder back bone to represent novel classes especially when the novel classes are significant in number. Direct fine-tuning of encoder backbone while training novel classes is not recommended as it affects the base class accuracy. We suggest utilising an incremental learning technique for GFSS to overcome this restriction. Our method avoids impacting base class accuracy with a specially designed Low Rank Adaptation (LoRA) based weight merging strategy.

The major contributions of our proposed work can be summarized as follows:

i) We introduce a novel incremental learning framework for learning the new classes without compromising on the base class accuracy by utilizing LoRA based fine-tuning framework
ii) We design a novel weight initialization strategy for LoRA based fine tuning suited for the proposed incremental learning frame work
iii) We also propose a novel weight merging strategy across various encoder layers considering different rates of 'knowledge assimilation' in encoder layers.

Roy, A. and Gopalakrishnan, V.
Weight Factorization Based Incremental Learning in Generalized Few Shot Segmentation.
DOI: 10.5220/0013187700003912

In *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2025) - Volume 3: VISAPP*, pages 556-562
ISBN: 978-989-758-728-3; ISSN: 2184-4321

iv) Finally, we adapt the novel class prototype learning to our proposed incremental learning framework

2 PAST WORKS

The works on few-shot semantic segmentation (FSS) (Shaban et al., 2017) have gained much interest after the success of several few-shot learning (Finn et al., 2017)(Ravi and Larochelle, 2017) methods. Few-shot Semantic Segmentation uses just a few available annotated support images to provide a dense pixel level label prediction for the query images. FSS techniques(Dong and Xing, 2018)(Rakelly et al., 2018) primarily relied on a dual-branch architecture, in which a query image is segmented using learned prototypes from support images. Many of the later works investigated on various sophisticated methods to assimilate knowledge from support samples to improve query image segmentation (Li et al., 2021)(Liu et al., 2022b)(Tian et al., 2022b) (Wang et al., 2019)(Yang et al., 2020)(Zhang et al., 2021). These methods included building multiple prototypes per class each activating different regions of the query image, using Graph CNNs to establish correspondences between support and query images (Wang et al., 2020), imprinting weights for new classes (Siam et al., 2019), and utilising vision transformers to better transfer category information (Liu et al., 2022a)(Lu et al., 2021)(Zhang et al., 2022).

Even with significant work done on FSS, the simultaneous segmentation of both base and novel classes in an image remained a challenge for few-shot segmentation techniques. To address this issue, Tian et al. developed the Generalised Few-shot Semantic Segmentation (GFSS) technique in their pioneering work(Tian et al., 2022a). Their method segments both known and new object classes from an image in one step, without needing paired support images. It uses contextual information from the support and query images to improve the classifier for segmenting the new object classes. Some other recent works in GFSS include: (Liu et al., 2023b) proposing graph network-based class contrastive loss minimizing intra-class variations, maximizing inter-class dissimilarity; (Hajimiri et al., 2023) using DIaM to maximize mutual information between features-predictions while ensuring consistency with prior model via KL divergence; (Huang et al., 2023) employing foreground perception module, kernel techniques for observed classes, prototype learning for novel objects; (Lu et al., 2023) using Transformer-based calibration module balancing base-novel class predictions,

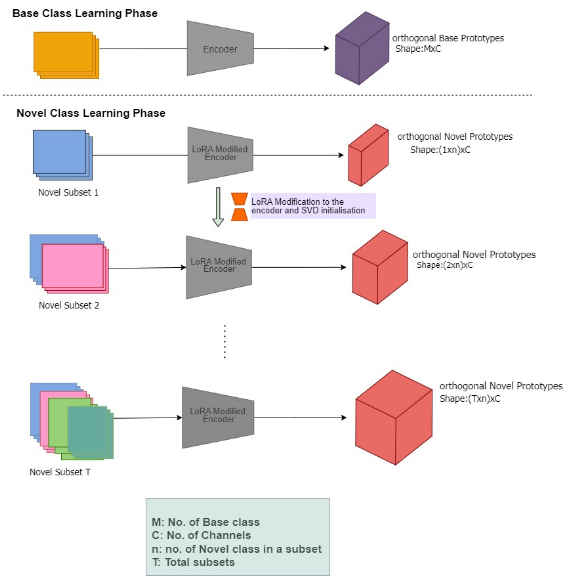

Figure 1: Training framework of proposed system. Novel classes are divided to T subsets and incrementally learnt using LoRA based backbone adaptation.

trained episodically using cross-covariance between features-scores.

The current state-of-the-art GFSSeg method (Liu et al., 2023a) builds a set of orthogonal prototypes, each of which represents a semantic class, and makes predictions for each class independently based on the features projected onto its prototype. This technique builds upon all of the seminal work on GFSS. Moreover, by employing a residual of feature projection as the background representation, it tackles semantic shifting, which occurs when the backdrop no longer contains pixels from novel classes during the updating phase.

3 PROPOSED WORK

3.1 Few Shot Segmentation with Orthogonal Prototypes

Our proposed method named as WFIL (Weight Factorization based Incremental Learning), tries to overcome the limitations of novel class learning in prior works by incrementally absorbing the additional information regarding novel categories into the neural network architecture learnt for base classes. Though our proposed strategy is generic enough to be incorporated to different GFSS architectures, we use the architecture in (Liu et al., 2023a) as a baseline to our experiments.(Liu et al., 2023a) learns orthogonal class prototypes for novel categories using specially designed orthogonal loss functions and achieves state-of-the-art accuracies in genaralized few shot segmentation. With an objective to retain the base class accuracy while learning novel classes, (Liu et al., 2023a) freezes the encoder backbone. We argue that freezing the encoder backbone can potentially limit the flexibility of learning novel classes as explained before. As a better alternative , we propose an incremental learning mechanism for novel classes wherein we can achieve improved segmentation accuracies for novel classes while retaining most of the accuracy of base classes.

3.2 Incremental Learning in Few Shot Segmentation

As mentioned earlier, the baseline architecture in (Liu et al., 2023a) learns orthogonal prototypes for novel categories but retains the same encoder backbone to maintain base class accuracy. We need to adapt the encoder backbone to efficiently incorporate the novel information but should be careful enough not to unlearn the information learnt for base classes to retain base class segmentation accuracy. Low Rank Adaptation (LoRA) method (Hu et al., 2021) originally proposed for efficient fine-tuning in Large Language Models (LLMs) can be effectively adapted to suit our requirements in GFSS learning as well. The key idea behind LoRA is that while the weight matrices in a deep learning model's dense layers are usually full-rank, when the model adapts to a particular task, the pre-trained models demonstrate a low "intrinsic dimension". Hence the models can still learn effectively despite projection to a smaller subspace. In this work, for the first time, we comprehensively explore the influence of modifying the encoder weights during the GFSS novel class training phase. As LoRA fine tuning maintains the original weights frozen, the target of maintaining base class accuracy will be also achieved. In LoRA (Hu et al., 2021), a low-rank decomposition of the original weight matrix $\mathbf{W}_0$ is performed using $\mathbf{W}_0 + \Delta\mathbf{W} = \mathbf{W}_0 + \mathbf{BA}$, where $\mathbf{B} \in \mathbb{R}^{d\times r}$, $\mathbf{A} \in \mathbb{R}^{r\times k}$, and the rank $r = \min(d,k)$, is used to confine the update of any pre-trained weight matrix $\mathbf{W}_0 \in \mathbb{R}^{d\times k}$. While $\mathbf{A}$ and $\mathbf{B}$ have trainable parameters, $\mathbf{W}_0$ is fixed and does not receive gradient changes throughout training. $\mathbf{W}_0$ and $\Delta\mathbf{W} = \mathbf{BA}$ are multiplied with the same input, and their respective output vectors are summed coordinate-wise. After the LoRA change, the forward pass that previously looked like $\mathbf{h} = \mathbf{W}_0\mathbf{x}$ gives $\mathbf{h} = \mathbf{W}_0\mathbf{x} + \Delta\mathbf{W}\mathbf{x} = \mathbf{W}_0\mathbf{x} + \mathbf{BAx}$

For the efficient learning of the LoRA parameters, we adopt an incremental way of training the encoder backbone. In the GFSS setting we have a certain dataset containing the base classes set C_{base} with adequate data available and novel classes set C_{novel}, with a small amount of annotated data. We divide the C_{novel} into subsets $\mathcal{D} = \{(X_i,Y_i)\}_{i\in\{1,2,\ldots,m\}}$, where m is the total number of classes in that particular subset, X_i is the input image for the i-th class, and Y_i is the corresponding ground truth segmentation mask for the i-th class. To progressively add additional classes, a subset $\mathcal{D}_t$ contains all the classes from $\mathcal{D}_{t-1}$ along with k new classes that were previously not there in the subset $\mathcal{D}_{t-1}$.Each subset $\mathcal{D}_t$ is made up of images and their ground truth mask corresponding to a set of classes C_t, such that $C_1 \cup C_2 \cup \ldots \cup C_T = C_{\text{novel}}$. During the incremental learning process, the model is trained and evaluated iteratively on one subset at a time; hence, if there are T subsets, the model will be trained for T iterations. This method of incremental training has significantly improved the baseline model.

3.3 Weight Factorization Based Initialization for Incremental Learning

In LoRA based fine tuning (Hu et al., 2021), the new trainable weight matrices A and B are initialized with a random Gaussian initialization and with zeros respectively. As our objective is to learn incrementally the new information from novel class set, we propose to initialize the weight matrices A and B with 'condensed knowledge' of the past learnt information. Hence we initialize the $lora\_A$ and $lora\_B$ weight matrices with the top r entries in the singular value decomposition of the original weight matrix W_0, as shown in Figure 2. Figure 2(a) shows the initialization from (Hu et al., 2021) and figure 2(b) shows our modified initialization for incremental GFSS learning. The SVD re-constructed value of W_0, which comprises the weights from the base class learning phase, is used to initialize the $lora\_A$ and $lora\_B$ weight matrices during the first iteration of the novel class learning phase.In subsequent iterations, the weight matrices for $lora\_A$ and $lora\_B$ are initialized using SVD decomposition of the weight matrix from the previous iteration, thereby transferring the knowledge learned in iteration $T-1$ to the current iteration T, ensuring continuity in the learning process.

3.4 Rank Based Layer-Wise Scaling and Weight Merging

According to (Hu et al., 2021), the rank r is a hyperparameter that controls the size of the low-rank ma-

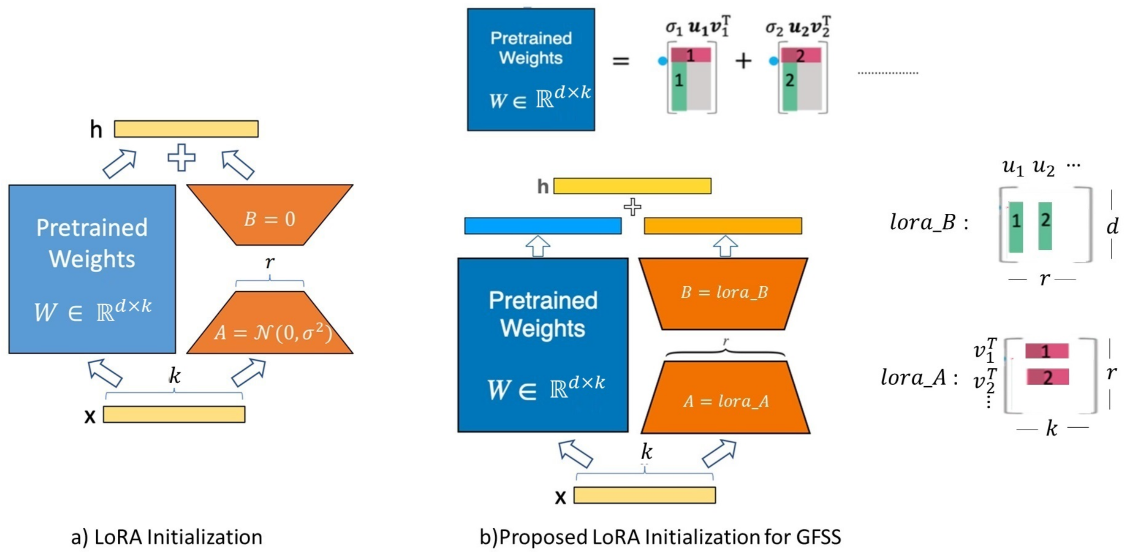

Figure 2: SVD based initialization in LoRA based incremental learning for GFSS.

trices used to update the pre-trained model's weights. In our method, the new weight values are initialized with the SVD decomposition of W_0 considering **top** r singular vectors, where r is decided by fixing a pre-defined percentage drop in **top** singular value of W_0. We name the selected value r as the 'rank of incremental update'. For instance, we retain only those singular values that are 70 percent or more of the top singular value. This threshold, however, is specific to the dataset being used. This approach aims to retain the most significant components of the weight matrix while ignoring insignificant noise. The LoRA parameters and the initial weight W_0 are multiplied with the same input during the model assessment phase, and the corresponding output vectors are then added coordinate-wise:

$$Wx + \Delta W = W_0 x + \delta . BAx = h \tag{1}$$

A scaling factor δ is used to scale down the LoRA weights while adding them back to the original weights.This regulates how much the original weights are updated from LoRA based training. We propose to use different scaling factors to different layers in the encoder backbone as the later will need different amounts of incremental update across the layers. Our studies indicate that rank of the original weight matrix increases differently across every layer in each iteration of incremental learning. This demonstrates the capacity of the model to learn from the data presented differently across various layers. The incremental update needed for each encoder layer in n^{th} iteration can be related to the delta change in the ranks of $(n-1)^{th}$ and n^{th} iterations for the same layer. Scaling for layer l is, $\delta = \frac{1}{1+\delta_l}$, where δ_l is the rank difference between the $(n-1)^{th}$ and n^{th} iterations for l^{th} layer weight matrix. A larger rank difference in subsequent iterations for a layer implies a larger displacement of weights from the base encoder. To maintain the base accuracy, we inversely scale the LoRA weights based on the incremental rank update from each layer.

3.5 Incremental Update of Novel Class Weight Prototypes

In our proposed method of incrementally learning orthogonal prototypes to represent novel classes, the set of all novel classes is partitioned into subsets as mentioned earlier. The learning process proceeds iteratively over these subsets. For the initial subset, a prototype weight matrix is learned with orthogonal initialization .The weight matrix has a shape of n x C, where n is the number of classes in the subset, and C is the number of channels. In subsequent iterations, when a new subset of classes is introduced, the prototype weight matrix from the previous iteration is augmented with additional rows initialized with orthogonal weights to accommodate the new classes. The augmented weight matrix is then fine-tuned learning the orthogonal representation of the new classes in the current subset, simultaneously optimizing the previously learned weights representing the classes from the previous subset.This process continues iteratively, with the weight matrix being incrementally expanded and fine-tuned until all the novel classes have been covered. The final weight matrix represents the entire set of novel classes, leveraging the previously learned representations to facilitate incremental learning and potentially improving generalization and convergence compared to learning each subset independently from scratch.

4 EXPERIMENTS

We conducted experiments on the GFSS databases - Pascal-5^i and COCO-20^i to test our proposed method. As per the standard protocol of GFSS learning, the new classes in both datasets are evenly divided into four folds for cross-validation. Nevertheless, in accordance with our incremental learning approach, each fold's novel classes are separated into a number of subsets, with subset $\mathcal{D}_t$ including all the novel classes from subset $\mathcal{D}_{t-1}$ in addition to a few more new classes. Following (Liu et al., 2023a) once a model is validated on one fold, the classes in that fold serve as "novel classes," while classes from the other 3 folds plus background act as "base classes." In the base class learning phase, images containing at least one base class pixel are selected from the original training set. Novel class pixels in these images are treated as background during this stage. In our method, while incrementally learning novel classes in t^{th} iteration from a subset $\mathcal{D}_t$, we mimic a K-shot learning setting. K images for each of the novel classes included subset $\mathcal{D}_t$ are randomly sampled during the novel class updating phase. For instance, subset $\mathcal{D}_t$ may contain the three novel categories (c_1, c_2, and c_3). For subset $\mathcal{D}_t$, the K samples from each of these classes (c_1, c_2, and c_3) make up the few-shot training data. The training data for subset $\mathcal{D}_{t+1}$ comprises of the novel classes from subset $\mathcal{D}_t$ as well as three more novel classes that are absent from subset $\mathcal{D}_t$. The whole validation set is used to evaluate performance on both the base and novel classes. The model is optimised by utilising the mini-batch stochastic gradient descent with momentum of 0.9 and weight decay of 0.0001. The starting learning rate is set to 0.01 during base class learning, and it is annealed down to zero.

4.1 Training Strategy

Our proposed method builds upon the backbone encoder architecture introduced by Liu et al. (Liu et al., 2023a), which also incorporates the learning of orthogonal novel class prototypes. However, we diverge from their approach in a crucial aspect: the base-novel class split ratio. While Liu et al. used a 15-5 base-novel split for the Pascal-5^i database and a 60-20 split for the COCO-20^i database, we intentionally modify these ratios to challenge our method under more demanding conditions. Specifically, we employ a 5-15 base-novel split for Pascal-5^i and maintain the 60-20 split for COCO-20^i. This configuration for Pascal-5^i creates a scenario where novel classes significantly outnumber base classes. We kept the base-novel split for COCO-20^i unchanged because here the novel classes are already significant in number and owing to the challenges of the COCO-20^i dataset current models in GFSS struggle with the COCO-20^i dataset.Such a setup represents a more challenging task that current GFSS methods often struggle to address effectively. By testing our method under these conditions, we aim to validate its robustness and efficacy in scenarios with a high number of novel classes, thus pushing the boundaries of GFSS capabilities.

We have incrementally added novel classes to the training dataset until all novel classes are incorporated during the novel class updating phase. We have experimented with two different methods of incremental addition for Pascal-5^i Dataset: the first method included adding three novel classes to a subset $\mathcal{D}_t$ with each iteration, while the second method added five novel classes to a subset $\mathcal{D}_t$ with each iteration. For COCO-20^i Dataset we experimented by adding five additional novel classes to the subset $\mathcal{D}_t$ in each subsequent iteration using the incremental addition approach. Once the model is trained and evaluated on subset $\mathcal{D}_t$, for the training of the next subset $\mathcal{D}_{t+1}$, we again perform a LoRA decomposition of the latest encoder weights along with SVD initialization as mentioned in Section 3.3. For each subset we have trained for 200 epochs with a batch size of 2 for Pascal-5^i and for 500 epochs with a batch size of 8 for COCO-20^i dataset.

4.2 Result Discussion

All results presented in this paper are for one-shot segmentation (K = 1), which is most challenging task in GFSS. Table 1 shows the comparison between our proposed method and the state of the art method in GFSS , POP (Liu et al., 2023a). We conducted experiments for a split of 5 base classes and 15 novel classes for Pascal-5^i Dataset with two different incremental addition, 3 classes and 5 classes in every step. It can be observed that the proposed method achieves best accuracy numbers in base and novel classes when novel classes outnumber base classes.

We performed ablation studies to understand the impact of the different components in our proposed framework. Experiments are first performed with only incremental learning (IL) added to the POP architecture, as discussed in Section 3.5 . The encoder backbone is then modified with LoRA based fine-tuning using standard initialization (LoRA). Further the experiments are done with our proposed initial-

Table 1: Results for the Pascal-5^i Dataset with 5-15 Base-Novel split and incremental addition of 3 and 5 classes. We report mean Intersection over Union (mIoU) in percentage (%) across three categories: base classes (Base), novel classes (Novel), and all classes combined (Base + Novel = Total).

	Base (%)	Novel (%)	Total(%)
POP(Liu et al., 2023a)	51.98	23.51	31.65
WFIL (5 classes increment), ours	57.14	**29.50**	**37.40**
WFIL (3 classes increment), ours	**61.77**	27.48	37.27

Table 2: Ablation study on various modules in our method. The experiments are performed on Pascal-5^i dataset for a base-novel split of 5-15.

	5 classes increment			3 classes increment		
	Base	Novel	Total	Base	Novel	Total
IL	56.83	25.84	34.69	53.15	28.41	35.48
IL+LoRA	56.73	28.75	36.75	58.21	28.75	37.16
IL+LoRA+SVD	56.72	29.09	36.98	61.89	27.27	37.16
IL+LoRA+SVD+Scaling	**57.14**	**29.50**	**37.40**	**61.77**	**27.48**	**37.27**

ization (SVD) along with LoRA fine-tuning. Finally the rank adaptation based weight merging (Scaling) is applied to various layers in the encoder backbone. Table 2 captures accuracy values obtained for 5 class increment and 3 class increment. The improvement in accuracy numbers are consistent and substantiates the significance of each module proposed in our work.

Table 3 presents a comparative analysis of various GFSS methods, including our proposed approach, on the challenging COCO-20^i dataset. While most existing methods struggle with COCO-20^i dataset's inherent complexities, our method demonstrates superior performance over SOTA approaches. We maintain the conventional split of 60 base and 20 novel classes, as the number of novel classes is already significant and better represents real-world scenarios. The incremental learning process, which involves progressively adding five new classes to each subsequent subset, further validates our method's robustness. Notably, our approach achieves substantial improvements in novel class accuracy while effectively maintaining base class performance, demonstrating its capability to handle the complex nature of the COCO-20^i dataset, where current methods typically struggle to maintain such balanced performance.

These results align with our expectations, in our proposed model framework the modified version of LoRA fine-tuning is designed to enhance novel class accuracy while preserving base class performance. It's important to note that we do not present similar results for the traditional 15-5 base-novel split on the Pascal-5^i database. This is because our proposed method specifically targets scenarios with a high number of novel classes, which is not represented in the conventional Pascal-5^i split.

The experimental results demonstrated remarkable performance patterns across datasets. On Pascal, our method achieved significant improvements in both base and novel class accuracies, showcasing particularly strong performance on the challenging 5-15 Base-Novel split, which is notably more difficult than the traditional 15-5 split. This achievement is especially significant as the model effectively learned to perform well with fewer base classes while generalizing to a larger number of novel classes. For the more complex COCO-20^i dataset, where existing approaches typically struggle, our method not only maintained strong base accuracy but also achieved substantial improvements in novel class performance, surpassing previous SOTA results. This exceptional performance on COCO-20^i can be attributed to our method's robust learning framework, which effectively handles the dataset's inherent complexities including greater object variety, scale variation, and occlusion. The strength of our approach is validated by these close to real-world scenarios where data is presented with occlusion and scale variability and other challenges and a more challenging distribution that better reflect real-world scenarios where novel categories often match or exceed the number of base classes in quantity. These results demonstrate our method's superior capability in handling challenging few-shot semantic segmentation scenarios.

5 CONCLUSION

In this work we have showcased that the novel class accuracy in Generalized Few-shot Segmentation can be improved by a good margin if we adapt the encoder backbone using incremental learning. It was also shown that the base class accuracy can be retained when the backbone adaptation is done using LoRA fine-tuning along with our proposed weight ini-

Table 3: Comparison of GFSS accuracy on COCO-20^i Dataset across various methods. Our method (WFIL) uses the final combination of IL+LoRA+SVD+Scaling.

Method	Base (%)	Novel (%)	Total (%)
CAPL (Tian et al., 2022a)	44.61	7.05	35.46
Harmonizing Base and Novel class (Liu et al., 2023b)	46.89	8.83	37.48
PKL and OFP for GFS-Seg (Huang et al., 2023)	46.36	11.04	37.71
POP (Liu et al., 2023a)	**54.78**	18.07	45.71
WFIL (Our Method)	54.51	**21.93**	**46.46**

tialization and weight merging techniques. Our proposed method achieved state of the art accuracy values in Generalized Few-shot Segmentation when the number of novel classes are large or there is diversity and challenges in the dataset.

REFERENCES

Dong, N. and Xing, E. P. (2018). Few-shot semantic segmentation with prototype learning. In *British Machine Vision Conference*.

Finn, C., Abbeel, P., and Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1126–1135. PMLR.

Hajimiri, S., Boudiaf, M., Ayed, I. B., and Dolz, J. (2023). A strong baseline for generalized few-shot semantic segmentation.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. (2021). Lora: Low-rank adaptation of large language models.

Huang, K., Wang, F., Xi, Y., and Gao, Y. (2023). Prototypical kernel learning and open-set foreground perception for generalized few-shot semantic segmentation.

Li, G., Jampani, V., Sevilla-Lara, L., Sun, D., Kim, J., and Kim, J. (2021). Adaptive prototype learning and allocation for few-shot segmentation.

Liu, S.-A., Zhang, Y., Qiu, Z., Xie, H., Zhang, Y., and Yao, T. (2023a). Learning orthogonal prototypes for generalized few-shot semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Liu, W., Wu, Z., Zhao, Y., Fang, Y., Foo, C.-S., Cheng, J., and Lin, G. (2023b). Harmonizing base and novel classes: A class-contrastive approach for generalized few-shot segmentation.

Liu, Y., Liu, N., Yao, X., and Han, J. (2022a). Intermediate prototype mining transformer for few-shot semantic segmentation.

Liu, Y., Zhang, X., Zhang, S., and He, X. (2022b). Part-aware prototype network for few-shot semantic segmentation.

Lu, Z., He, S., Li, D., Song, Y.-Z., and Xiang, T. (2023). Prediction calibration for generalized few-shot semantic segmentation. *IEEE Transactions on Image Processing*, 32:3311–3323.

Lu, Z., He, S., Zhu, X., Zhang, L., Song, Y.-Z., and Xiang, T. (2021). Simpler is better: Few-shot semantic segmentation with classifier weight transformer.

Rakelly, K., Shelhamer, E., Darrell, T., Efros, A., and Levine, S. (2018). Conditional networks for few-shot semantic segmentation.

Ravi, S. and Larochelle, H. (2017). Optimization as a model for few-shot learning. In *International Conference on Learning Representations*.

Shaban, A., Bansal, S., Liu, Z., Essa, I., and Boots, B. (2017). One-shot learning for semantic segmentation.

Siam, M., Oreshkin, B., and Jagersand, M. (2019). Adaptive masked proxies for few-shot segmentation.

Tian, Z., Lai, X., Jiang, L., Liu, S., Shu, M., Zhao, H., and Jia, J. (2022a). Generalized few-shot semantic segmentation.

Tian, Z., Zhao, H., Shu, M., Yang, Z., Li, R., and Jia, J. (2022b). Prior guided feature enrichment network for few-shot segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(2):1050–1065.

Wang, H., Zhang, X., Hu, Y., Yang, Y., Cao, X., and Zhen, X. (2020). Few-shot semantic segmentation with democratic attention networks. In *European Conference on Computer Vision*.

Wang, K., Liew, J. H., Zou, Y., Zhou, D., and Feng, J. (2019). Panet: Few-shot image semantic segmentation with prototype alignment. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9196–9205.

Yang, B., Liu, C., Li, B., Jiao, J., and Ye, Q. (2020). Prototype mixture models for few-shot semantic segmentation.

Zhang, B., Xiao, J., and Qin, T. (2021). Self-guided and cross-guided learning for few-shot segmentation.

Zhang, G., Kang, G., Yang, Y., and Wei, Y. (2022). Few-shot segmentation via cycle-consistent transformer.