



Domain generalization for semantic segmentation: a survey

Taki Hasan Rafi¹ · Ratul Mahjabin² · Emon Ghosh³ · Young-Woong Ko¹ · Jeong-Gun Lee¹

Accepted: 28 May 2024 / Published online: 12 August 2024
© The Author(s) 2024

Abstract

Deep neural networks (DNNs) have proven explicit contributions in making autonomous driving cars and related tasks such as semantic segmentation, motion tracking, object detection, sensor fusion, and planning. However, in challenging situations, DNNs are not generalizable because of the inherent domain shift due to the nature of training under the i.i.d. assumption. The goal of semantic segmentation is to preserve information from a given image into multiple meaningful categories for visual understanding. Particularly for semantic segmentation, pixel-wise annotation is extremely costly and not always feasible. Domain generalization for semantic segmentation aims to learn pixel-level semantic labels from multiple source domains and generalize to predict pixel-level semantic labels on multiple unseen target domains. In this survey, for the first time, we present a comprehensive review of DG for semantic segmentation. We present a comprehensive summary of recent works related to domain generalization in semantic segmentation, which establishes the importance of generalizing to new environments of segmentation models. Although domain adaptation has gained more attention in segmentation tasks than domain generalization, it is still worth unveiling new trends that are adopted from domain generalization methods in semantic segmentation. We cover most of the recent and dominant DG methods in the context of semantic segmentation and also provide some other related applications. We conclude this survey by highlighting the future directions in this area.

Keywords Domain generalization · Out-of-distribution generalization · Domain shift · Semantic segmentation

1 Introduction

Machine learning plays a crucial role in various applications such as face recognition, self-driving cars, and healthcare (Roth et al. 2018; Rao and Frtunikj 2018; Guo and Zhang 2019), where a model needs to provide predictive decisions on the basis of the knowledge it acquired in the training stage. Nevertheless, what will happen if an autonomous system trained in California roads is tested on New York roads? How about a machine learning model that has been trained to perform sentiment analysis using data from the USA that is being used to interpret the sentiment of the posts from the United Kingdom? Again, Can a

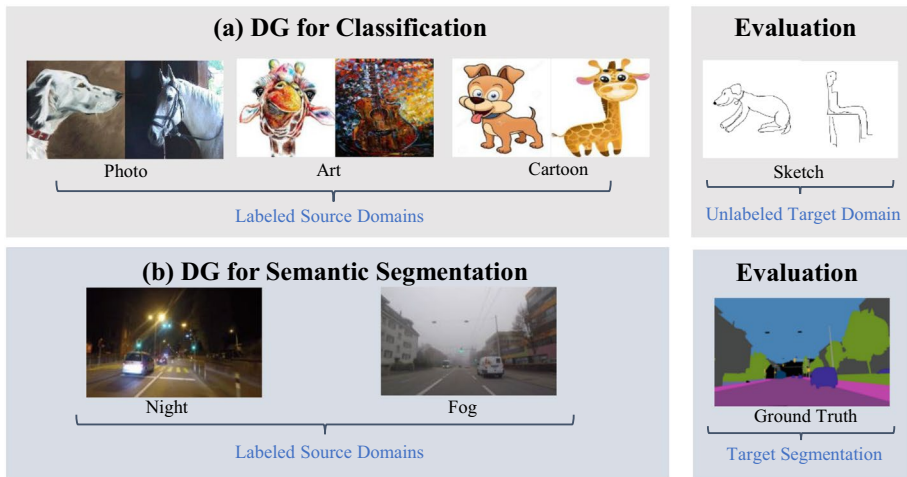


Fig. 1 Domain generalization for semantic segmentation. Examples from GTAV and ACDC datasets with different conditions such as Night and Fog for domain generalization, where the model is trained on the GTAV dataset, and directly evaluated in the ACDC dataset

tumor detection model trained on one group of patients work well for finding tumors in a diverse group with different health issues and backgrounds?

The answer to these questions is that those models will not work well based on the given situation. This arises from the assumption made by conventional machine learning (ML) techniques that the source and target data will be characterized by independence and uniformity. However, this assumption is not always fulfilled in reality. Often, data comes from different distributions introducing an issue known as *domain shift* (Ben-David et al. 2010; Blanchard et al. 2021; Moreno-Torres et al. 2012; Recht et al. 2019; Taori et al. 2020). Hence ML model experience notable decrease in performance when dealing with out-of-distribution (OOD) target domain (Fig. 1).

The issue of domain shift presents a substantial threat to the scalability of machine learning models across diverse applications within the field of Computer Vision. One such example is semantic segmentation (SS), a crucial computer vision task where each pixel in an image is categorized into a particular class (Guo et al. 2018). It has immense application for numerous applications such as autonomous driving, medical imaging, image editions, etc. Building a robust semantic segmentation model that can work well in unfamiliar situations (new unseen domains) is essential. A direct approach to address the domain shift challenge is to gather data from every possible domain which is both costly and practically unfeasible. Another alternative approach involves data collection from the target domain to adapt the trained model, referred to as *domain adaptation*, on the source domain. It is not always feasible in real-world scenarios (Wang et al. 2022a). In many cases, it is difficult to gather or even unknown before deploying the model, e.g., in Biomedical applications where it is impractical to collect new patient data in advance. So it is crucial to enhance the model's generalization capability. To address this issue without requiring data from the target domain, *domain generalization* (DG) was introduced (Muandet et al. 2013). The aim is to enhance the generalization capability of machine learning models by leveraging one or more related yet distinct source domains. In recent developments, domain generalization has been applied to advance the field of semantic segmentation. There are a few survey

papers that are based on domain generalization (Zhou et al. 2022a; Wang et al. 2022a), but they are not focused on a specific application of domain generalization but rather a general survey of domain generalization. This survey (Zhou et al. 2022a) mentioned the application of domain generalization in semantic segmentation. However, it does not hold a broad and comprehensive understanding of DG in semantic segmentation. Another survey (Li et al. 2023a) discussed the transformers for the segmentation task. There are several representative methods for semantic segmentation such as query-based and close-set segmentation methods (Cheng et al. 2020; Yu et al. 2018; Li et al. 2020b; Kirillov et al. 2020; Li et al. 2020b; Zhang et al. 2021a; Wang et al. 2021).

This paper presents the first comprehensive survey on Domain Generalization for Semantic Segmentation. We aim to introduce its recent advances, emphasizing its formulations, theories, algorithms, research areas, datasets, applications, and potential future research directions. We anticipate this survey will provide a comprehensive review for researchers interested in this topic and spark more research in this and related areas. There are several survey papers on domain generalization(DG) and semantic segmentation separately. However, to the best of our knowledge, this is the first paper addressing domain generalization in the context of semantic segmentation. Our contributions are summarized as follows.

- To the best of our knowledge, our survey is the first paper that comprehensively reviews domain generalization for semantic segmentation, which recently has caught growing attention in many computer vision applications.
- We discuss the widely used datasets, and evaluation metrics, and provide a quantitative comparison of the backbone segmentation models in different DG approaches.
- We provide future challenges and research directions that can be aggregated to solve underlying challenges in generalized semantic segmentation.

The rest of the paper proceeds as follows: Sect. 2 provides the necessary background. While Sect. 3 touches on related sub-areas. We explore various methodologies addressing Domain Generalization in Sect. 4 and discuss relevant datasets, benchmarks, and evaluation methods in Sect. 5. Section 6 houses a broad discussion on the future research directions. The paper concludes with Sect. 7.

2 Background

2.1 Problem formulation

Domain generalization or OOD generalization refers to signify the generalization capability on unseen target domains, and also in source domains. Here the target domains are denoted as $\mathcal{T} = \{\mathcal{T}_1, \dots, \mathcal{T}_N\}$. Usually, there are multiple sources $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_K\}$ to train and learn invariant semantic features. A semantic segmentation model ϕ outputs pixel-wise predictions p for given an image x . This semantic segmentation model consists of a feature extractor ϕ_{ext} and classifier ϕ_{cls} . While training the segmentation network, we have access to multiple source domains $\mathcal{D}_s = \{(x^s, y^s)\}$. Where, \mathcal{D}_s have multiple source domains \mathcal{S} . Here, each sample $x^s \in \mathbb{R}^{H \times W \times 3}$, corresponding pixel-wise labels $y^s \in \mathbb{R}^{H \times W \times 3}$. The segmentation loss for the baseline network ϕ can be calculated as,

$$\mathcal{L}_{ss} = -\frac{1}{HW} \sum_{h,w,k=1}^{H,W,K} y^s \log(\phi(x^s)) \quad (1)$$

The main goal is to minimize the source domain loss to ensure high generalization on unseen target domains \mathcal{T} , each \mathcal{T} is an unlabelled dataset $\mathcal{D}_t = \{(x^t)\}$. Traditionally, the segmentation model can be evaluated in both source \mathcal{S} and unseen target domains \mathcal{T} .

3 Deep learning methods in semantic segmentation: CNN and transformers

Recently, deep learning-based methods played an important role in semantic segmentation tasks. Semantic segmentation is also referred to as visual understanding. In recent times, CNN and vision transformer-based methods have mostly been used to solve challenges in semantic segmentation. In this section, we review some of the recent methods of CNN and vision transformer for semantic segmentation. DeconvNet (Noh et al. 2015) represents a significant contribution to this field, offering an approach that complements the conventional Fully Convolutional Network (FCN)-based methodologies, known for their proficiency in extracting a generalized form of objects. Contrasting the FCN, DeconvNet systematically organizes proposals by size, efficiently capturing multi-scale objects and discerning finer object details. Notably, the innovation of SegNet (Badrinarayanan et al. 2015, 2017) lies in its novel approach to upsample feature maps with low spatial dimensions within the decoder. Furthermore, SegNet incorporates a mechanism to retain the max-pooling indexes from the encoder's feature maps, bolstering its overall performance. Reference (Kendall et al. 2015) introduces a pixel-based probabilistic framework termed Bayesian SegNet, achieved by adapting the architecture of SegNet. This adaptation involves the implementation of a probabilistic encoder-decoder architecture using dropout (Srivastava et al. 2014), a technique also utilized in approximate inference by Bayesian CNN (Gal and Ghahramani 2015). Besides CNN, transformers are also extensively used in semantic segmentation. Strudel et al. (2021) used a vision transformer for semantic segmentation, they utilized output embeddings corresponding to image patches and used class labels from the embeddings with a mask transformer. There are other methods that use transformers (Xie et al. 2021; Zheng et al. 2021; Zhang et al. 2022b) for semantic segmentation.

4 Sub-related topics

The connections and differences between the DG on Semantic Segmentation and its related topics are addressed in this section.

Domain adaptation (DA) is an approach for improving a model's performance on a target domain with insufficient annotated data by using the information that the model has learned from a corresponding domain with enough labeled data. The goal of domain adaptation is to lessen disparities in the feature space across domains, both marginal and conditional. This involves identifying common underlying attributes shared involving the source and destination domains and subsequently adjusting them to enhance alignment. In other words, domain adaptation tries to lessen the detrimental impacts of domain shift, which can result in a decline in model performance for semantic segmentation when the model is applied to data from a distinct distribution. It is the subject that is most relevant to DG

and has been widely researched in the literature. Wang et al. (2023) Yang et al. (2022c) Xie et al. (2023) Toldo et al. (2022) Shyam et al. (2022b). Wang et al. (2023) proposed a target-to-source DA technique to encourage the model to learn comparable cross-domain properties using a dynamic re-weighting strategy to help the model. Yang et al. (2022c) gave the idea of a framework that is easy to train and learns domain-invariant prototypes for domain adaptive semantic segmentation. In order to encourage the learning of class-discriminative and class-balanced pixel representations across domains and ultimately improve the performance of self-training methods, Xie et al. (2023) introduced an innovative concept named Semantic-Guided Pixel Contrast (SePiCo). A unique one-stage adaptation model that emphasizes the semantic ideas contained in each individual pixel. Toldo et al. (2022) introduced that when learning incremental tasks, style transfer strategies are used to expand knowledge between domains, and a strong distillation framework is used to successfully remember task information under incremental domain change. In order to improve an underlying segmentation network such that it consistently performs in unidentified actual destination domains, Shyam et al. (2022b) proposed the notion of utilizing a large number of synthetic source domains. Yang et al. (2023b) recommended a unique Sparse Visual Domain Prompts (SVDP) method in order to overcome domain shift issues in semantic segmentation. It realizes effective cross-domain learning and seeks to extract more regional domain-specific information. Self-ensembling models, provide a different perspective on how to learn domain-invariant properties and introduce domain adaptability for semantic segmentation (Xu et al. 2019). He et al. (2021) offered an interactive learning method for domain adaptation without investigating any data from the desired domain to make full use of the vital semantic knowledge across source domains. The basic goal of domain generalization is to develop a model that can perform well on destination domains that were not encountered during training. The objective is to generalize across multiple domains. The similarities between DG and DA on semantic segmentation are the existence of domain shifting and the transfer of knowledge between source and destination domains. On the contrary, DG deals with an unseen target domain and DA addresses a known target domain. There are different settings proposed such as source-free DA, where source-domains are not utilized during the testing. Liu et al. (2021b) proposed a distillation-based source-free domain adaptation method that preserves the source knowledge via knowledge transfer to retain contextual feature relationships for semantic segmentation. Yang et al. (2022a) suggested a source-free domain adaptation method that is based on self-training and distribution transfer by aligning implicit feature representation of the source model. Kundu et al. (2021) proposed a source-free DA method based on pseudo labeling that is generated by a multi-head framework. On top of it, they proposed a conditional prior-enforcing auto-encoder to retain high-quality pseudo labels in the target domain. You et al. proposed a source-free DA method based on positive and negative learning, where the main mechanism is to select class-balanced pseudo-labeled pixels, where negative learning does the heuristic complementary label selection. Some other related DA methods such as unsupervised DA (Zou et al. 2018; Zhang et al. 2019; Lee et al. 2021; Sankaranarayanan et al. 2017), semi-supervised DA (Chen et al. 2021b; Wang et al. 2020b; Hoyer et al. 2023), and few-shot DA (Kalluri and Chandraker 2022; Lei et al. 2022) also solved the similar semantic segmentation problem.

Self-supervised learning (SSL) aims to tackle problems by pretraining a general model with an enormous quantity of unlabeled data and subsequently tuning it on a downstream task with a limited amount of labeled data (Ziegler and Asano 2022). Its effectiveness is its capacity to make use of enormous quantities of unlabeled data and build accurate representations that highlight certain patterns and structures in the data. It can be used to pre-train

models and learn general-purpose representations, which can be transferable and useful for domain generalization. The choice of the self-supervised task and the domain difference among the pretraining and evaluation datasets play a crucial factor in determining the model's success in SSL. It seeks to enhance performance on a particular task, such as semantic segmentation, by utilizing unlabeled data through auxiliary tasks. On the other hand, domain generalization focuses on making the model robust to handle different and unseen data distributions, enabling it to perform well in diverse real-world scenarios.

Semi-Supervised Learning (SeSL) is a branch of machine learning that emphasizes carrying out particular learning tasks using both labeled and unlabeled data (Van Engelen and Hoos 2020). The segmentation performance is further enhanced by including prediction filtering into the already established SWSS algorithms (Bae et al. 2022). A semi-supervised framework built on Generative Adversarial Networks (GANs) was suggested by Souly et al. (2017) to ensure improved quality of images for GANs and subsequently better pixel classification. These approaches were tested on various challenging comparative visual datasets, i.e. PASCAL, SiftFlow, Stanford, and CamVid. A boundary-optimized co-training (BECO) method has been implemented to train the segmentation model in consideration of the noisy pseudo-labels, and WSSS should be converted to robust learning Rong et al. (2023). Kweon et al. (2023) proposed a completely new WSSS framework via adversarial learning of a classifier and an image reconstructor. To address the noise label and multi-class generalization issues, Chen et al. (2023a) suggested an end-to-end multi-granularity noise reduction and bidirectional alignment (MDBA) model. With simple-to-complex picture synthesis and complex-to-simple adversarial learning, this approach is suggested to close the data distribution difference in both input and outcome space. An integrated transformer architecture was proposed by Lian et al. for learning two modalities of class-specific tokens, i.e., class-specific visual and textual tokens. In semantic segmentation, domain generalization involves the procedure of training a model to perform well on semantic segmentation tasks across various source domains to improve its ability to generalize well to an unknown destination domain. The primary distinction between SSL and DG is that semi-supervised learning often assumes that the unlabeled data comes from exactly the identical distribution as the labeled data.

Multi-Task Learning (MLT) is a machine learning paradigm that attempts to capitalize on valuable knowledge from a variety of associated tasks to enhance the generalization efficiency of all the tasks (Zhang and Yang 2021). While DG aims to generalize a model to unknown data distribution, MTL aims to improve a model's performance on the exact same set of tasks that the model was trained on. Using knowledge obtained from numerous diverse independent data sources, Graham et al. (2023) proposed a multi-task learning method for segmenting and categorizing nuclei, glands, lumina, and various tissue regions. Bischke et al. (2019) dealt with the issue of maintaining semantic segmentation borders in high-resolution satellite imagery by using a recent multi-task loss methodology. The bias resulting from the loss causes the network to give greater attention to pixels close to boundaries by using several output descriptions of the segmentation mask. Semantic segmentation performance is improved by multi-task self-supervised learning with no additional annotation or inference-related computing costs (Novosel et al. 2019). Lu et al. (2020) suggested model can learn segmentation and per-pixel depth regression from a single image input by using multi-task learning. Researchers introduced a novel approach to simultaneously estimate disparity maps and segment images by combining the training of an encoder-decoder-based interactive convolutional neural network (CNN) for single image estimation of depth and a multiple class CNN for image segmentation. In order to improve the super-resolution model toward generating images that are most appropriate for the

purpose of segmentation rather than ones that are of high fidelity, Aakerberg et al. (2021) introduced an approach that jointly trains a high-resolution and semantic segmentation model from beginning to end manner using exactly the same task loss for both models. In parallel, researchers updated the segmentation model to more effectively use the enhanced images and raise segmentation accuracy. An innovative multi-task learning technique for the categorization of tumors in ABUS images implementing an encoder-decoder network and a lightweight multi-scale network has been developed (Zhou et al. 2021). A new sharing unit called a cross-stitch unit, that can be trained end-to-end, combines the activations from several networks (Misra et al. 2016). The goal of multi-task learning for semantic segmentation is to jointly build a model to carry out a variety of segmentation-related tasks by utilizing shared representations. By employing data to train a model from many source domains, domain generalization for semantic segmentation tries to make a model resilient to domain transformation and enable it to function well on an unknown destination domain. In order to enhance the performance of the model, both strategies use shared knowledge, but they focus on different problems: task diversity in multi-task learning and domain shift in domain generalization.

Transfer Learning (TL) focuses on transferring knowledge from one (or more) problem/domain/task to another but associated one (Pan and Yang 2009). Fine-tuning is a widely recognized example in contemporary deep learning: pre-train deep neural networks on enormous datasets, such as ImageNet (Deng et al. 2009) for vision models or BooksCorpus (Zhu et al. 2015) for language models, and then improve them on subsequent tasks (Girshick et al. 2014). To bridge the gap between the large source domain and the constrained destination domain, Sun et al. (2019) suggested a technique that makes use of transfer learning for semantic segmentation. It adapts to the destination domain using both actual and synthetic images as learning sources. Without taking into account the supine or prone positions, Ham et al. (2023) suggested that Semantic Segmentation uses transfer learning of convolutional neural networks to perform robust breast segmentation in supine breast MRI. Yang et al. (2021) suggested an effective semantic segmentation technique that makes use of the feature extractor of a real-time object detection model. Nigam et al. (2018) presented a new dataset and suggested a successful method for comparing train and test distributions with totally distinct scene organization, views, and object statistics. A common transfer learning strategy is pretraining-finetuning, in which the tasks for the source and destination domains are different and the destination domain can be accessed during training. The training and test tasks are typically the same despite having different distributions, and the target domain is not available in DG. In contrast to DG, which assumes no access to the target data and instead focuses on model generalization, TL requires the target data for model fine-tuning for new downstream tasks.

Few-Shot Meta-Learning (FSML) is a machine learning technique that uses a minimal number of labeled samples per class to guarantee that a pre-trained model generalizes across new types of data (that the pre-trained model has not seen in training). It is unique compared to traditional supervised learning. Traditional supervised learning methods require a huge amount of labeled training data. The test set also contains samples of data that must have a similar statistical distribution and come from the same categories as the training set. However in the case of FSML, even if the model was pre-trained using a statistically different distribution of data, the model can be used to expand to additional data domains as long as the data in the support and query sets are coherent. Pambala et al. (2021) proposed Semantic MetaLearning (SML), a modern meta-learning system that builds prototypes for a select group of annotated training images that includes class-level semantic descriptions. Tian et al. (2020) introduced the MetaSegNet framework for

multi-object segmentation. In order to extract the appropriate meta-knowledge for the few-shot segmentation, an embedding module architecture composed of the global and local feature branches was developed. A novel Cycle-Resemblance Attention (CRA) module has been added to a special self-supervised few-shot medical image segmentation network in order to make full use of the pixel-wise relationship between the query and support medical pictures (Ding et al. 2023). In order to conquer the difficult CD-FSS problem, Lei et al. (2022) introduced a novel Pyramid-anchor-transformation-based few-shot segmentation network (PATNet) that converts domain-specific attributes into domain-agnostic ones for downstream segmentation modules to quickly adapt to unknown domains. For learning semantic alignment with query features, Chen et al. (2021a) presented a class-specific blueprint and a class-agnostic blueprint and produced complete sample pairs. Li et al. (2021c) proposed method produces arbitrary pseudo-classes at random in the background of the query photos, supplying additional training data that is not available when forecasting particular target classes. The objective of domain generalization is to improve the robustness and generalization ability of models across various domains by addressing domain shifts. The similarity between DG and FSML is that, both the strategy increase the generalization ability of models. But in the case of FSML, it focuses on adjustment on new tasks while in DG it enhances the competence of the model to perform well on unknown data distribution.

5 Methodology

There are three categories in domain generalization (Wang et al. 2022a), such as (a) Data Manipulation, where it manipulates the input for better learning of the data, e.g. data augmentation, generation, normalization, and randomization fall into this category. (b) Representation Learning, which is apparently the most popular category, e.g. Domain invariant feature representation and feature disentanglement, where features are disentangled for domain-specific learning, and lastly, (c) Learning Strategy, which focuses on general learning capabilities to improve generalization, e.g., meta-learning, self-supervised learning. As mentioned, these categories are also divided into sub-categories. In this section, we provide a detailed explanation of existing domain generalization (DG) methods for semantic segmentation (SS). Figure 2 depicts the structure of the categories of domain generalization.

5.1 Data augmentation

Augmentation techniques have been found in extensive use in supervised learning for training machine learning models to reduce overfitting problems by enhancing the generalization performance of a model (Honarvar Nazari and Kovashka 2020; Shorten and Khoshgoftaar 2019; Khosla and Saini 2020; Yang et al. 2022b). The fundamental concept involves augmenting the original pairs (x, y) with new pairs $(A(x), y)$, where $A(x)$ denotes the transformation applied to x . Invariably they can be adopted for DGSS. In their work, Xu et al. introduced a novel data augmentation strategy called “amplitude mix,” which relies on Fourier-based techniques. This method involves interpolating between the amplitude spectrums of two images in order to preserve phase information (Xu et al. 2021). Su et al. proposed *SLAug*, Saliency-balancing Location-scale Augmentation (LLA) comprising Global scale Augmentation (GLA) for increasing source-like images through global distribution shifting and LLA for conducting class-specific augmentation (Su et al. 2023). Inspired

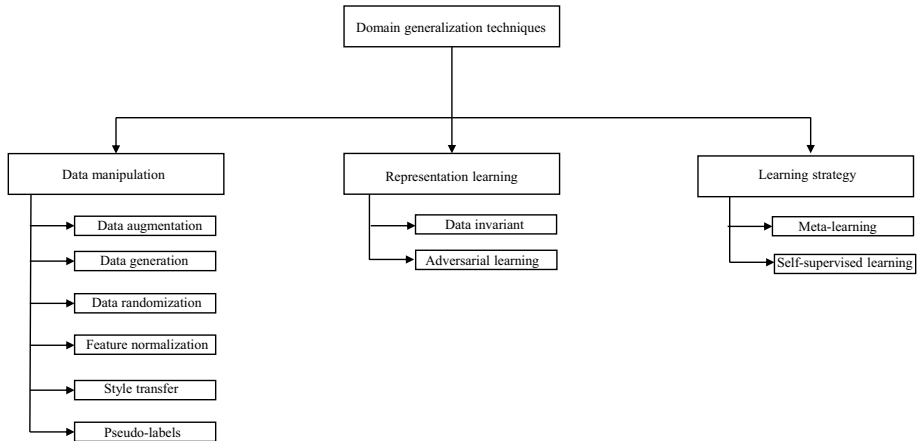


Fig. 2 Taxonomy of domain generalization methods

by topology-altering augmentation techniques (Chen et al. 2019; Dwibedi et al. 2017; Kumar Singh and Jae Lee 2017; Yun et al. 2019), Sellner et al. (2023) demonstrated *Organ Transplantation* to address geometric domain shifts based on application-specific data augmentation. Based on adversarial style augmentation, Zhong et al. (2022) introduced an innovative augmentation approach named *AdvStyle*. This technique generates challenging stylized images during the training process, effectively countering overfitting on the source domain. Kim et al. (2023a) proposed LiDAR semantic segmentation (DGLSS) by augmenting domains with diverse sparsity. Shyam et al. (2022a) introduced a style mixing augmentation that leads to features belonging to the same category having different styles. To address blind feature alignment, Shen et al. (2023) proposed a cross-domain mixture data augmentation technique. Zhao et al. (2022a) proposed a clustering instance mix (CINMix) augmentation technique to diversify the layouts of the source data. Lyu et al. (2022) introduced an approach called *Automated Augmentation for Domain Generalization (AADG)*. This work aimed to create novel domains through a proxy task to enhance diversity in the context of retinal image segmentation.

5.2 Domain randomization

Domain randomization (DR) is a technique for improving the generalization ability of ML models to new domains. This work involves the stochastic generation of synthetic data encompassing a wide range of potential domains. This contributes to learning domain invariant features such as lighting, object pose, and background clutter. Wu et al. (2022) proposed a “SiamDoGe” segmentation method that hinges upon a feature randomization technique with the objective of learning domain invariant features. Gong et al. (2022) formulated a strategy known as *Class Mixed Sampling Intermediate Domain Randomization (CIDR)* which works between source and pseudo-target domain. Peng et al. (2021) introduced Local Texture Randomization (LTR) and Global Texture Randomization (GTR) to induce randomization into the texture of source images for the diversification in texture styles. Xiao et al. (2023) designed PointDR that alternatively randomizes the geometry styles of the point clouds and aggregates their embeddings for the purpose of broadening training point cloud dataset distribution for 3D segmentation.

5.3 Domain generation

Data generation is a technique for improving the generalization of machine learning models to novel domains. This is achieved through the generation of synthetic data covering a diverse range of domains. Chen et al. (2023b) proposed a Generative Semantic Segmentation(GSS) model based on Vector Quantized Variation AutoEncoder(VQVAE). Li et al. (2021a) introduced an innovative generative framework built upon StyleGAN2 (Karras et al. 2020). It is tailored for addressing generative semantic segmentation tasks utilizing generative models with joint image-label distribution. Zhao et al. (2022b) proposed *Style-Hallucinated Dual Consistency learning*(SHADE) framework. It was introduced to address domain shift challenges in the context of semantic segmentation.

5.4 Domain adversarial learning

Domain adversarial learning can be used for semantic segmentation for learning domain invariant features. Ganin and Lempitsky (2015) first introduced *Domain-Adversarial Neural Network*(DANN) with the objective of adaptation between source and target domain. In the architecture, a single network accommodates both the generator and discriminator. The generator tries to fool the domain classifier and the domain classifier forces the generator to extract domain-invariant features. Tjio et al. (2022) proposed an adversarial semantic hallucination(ASH) approach with the aggregation of a class-conditioned hallucination module and a semantic segmentation module. Similar to the generator and discriminator, the segmentation module and hallucination module challenge each other to boost the generalization capability of the model. Xu et al. (2022a) proposed an adversarial framework for organ segmentation from a single domain to ensure semantic consistency through contrastive learning with Mutual information regularizer. To improve cooperation between domains, Zhang et al. (2023) introduced MTDA, a self-training method combining pseudo-labeling and feature stylization. Xu et al. (2022a) also proposed a novel adversarial DG method for organ segmentation trained on a single domain. A novel component *Adversarial Domain Synthesizer*(ADS) was incorporated to enable effective training on a single domain in the presence of domain shift. GAN-based method presented by Sankaranarayanan et al. (2018) to align the source and target data samples in the latent feature space.

5.5 Self supervised learning

Self Supervised Learning (SSL) can be used to improve generalization. The key idea is that a model learns generic features regardless of the target task by solving pretext tasks. Without the need for any domain labels, it can be used for semantic segmentation in single and multi-source settings (Zhou et al. 2022a). Vertens et al. (2020) proposed a multimodal semantic segmentation model utilizing a teacher-student training approach that transfers knowledge from the daytime domain to the nighttime domain. Yang et al. (2023a) proposed a Domain Projection and Contrastive Learning(DPCL) approach including self-supervised domain projection(SSDP) and multi-level contrastive learning(MLCL). SSDP aims to lessen the domain gap by projecting to the source domain. Zhou et al. (2022b) presented a multi-task paradigm with *domain-specific image restoration*(DSIR) module employing self-supervision.

5.6 Meta learning

Meta-learning is referred to as “*learn to learn*”. It can quickly adapt to new tasks with limited data by learning from a variety of tasks. The goal of meta-learning is to use prior knowledge from the learned tasks to handle new tasks efficiently. Since it can be employed to increase generalization, it can be used for semantic segmentation tasks learning from a variety of complex scenarios. Kim et al. (2022) presented a memory-guided domain generalization method based on a learning framework. Zhang et al. (2022a) introduced a novel domain for semantic segmentation that takes advantage of model-agnostic learning. Dou et al. (2019) adopted a model-agnostic learning paradigm with gradient-based meta-learning. They introduced a pair of complementary losses designed to effectively regulate the semantic structure of the feature space. Gong et al. (2021) proposed a meta-learning-based strategy for addressing Open Compound Domain Adaptation(OCDA) in the context of semantic segmentation. Shiau et al. (2021) addressed domain generalized semantic segmentation by proposing a novel meta-learning scheme with feature disentanglement ability. Zhang et al. (2022a) developed a domain generalization framework that jointly exploits the model-agnostic training scheme and target-specific normalization test strategy for semantic segmentation tasks. Qiao et al. (2020) introduced adversarial domain augmentation to counter the OOD generalization problem by leveraging the meta-learning framework.

5.7 Feature disentanglement

Feature disentanglement refers to the process of separating the factors of variation by breaking down the learned representations of the data. In the context of DG, it can be used to separate the domain-specific and domain-invariant features in the data. It focuses on the features that vary across domains by learning domain-invariant features. Jin et al. (2021) designed a Style Normalization and Restitution module(SNR) where disentanglement aims at better restitution. Bi et al. (2023) proposed a novel mutual information(MI) based framework to disentangle the anatomical and domain feature representations. Similar to this work (Bi et al. 2023), Li et al. (2021b) utilized MI-based disentanglement representation for left atrial(LA) segmentation.

5.8 Feature normalization

Feature normalization is a process to standardize data into uniform and stable distribution without extra data (Liu et al. 2023). Liu et al. (2023) proposed the spectral-spatial normalization(SS-Norm) module to enhance the generalization ability of the model. Bahmani et al. (2021) enhanced the inference procedure with normalization layers.

5.9 Domain invariant

The main objective of domain invariant representation-based approach is to learn domain invariant features from source(s) that will be applied to target as well. By leveraging general semantic shape priors, Liu et al. (2022) presented a novel approach *Test-time Adaptation from Shape Dictionary* (TASD) to overcome the single domain generalization problem for medical image segmentation. Xu et al. (2022) proposed

Domain-invariant Representation Learning(DIRL) algorithm to realize the quantification and utilization of the feature prior to urban-scene segmentation. Liao et al. (2023) introduced a domain generalization approach for semantic segmentation exploiting edge and semantic layout reconstruction to clarify content information. He et al. (2023) designed *Patch Statistical Perturbation* (PSP) to enhance the patch diversity facilitating the model to learn features that are domain invariant.

5.10 Pseudo label

Pseudo-label can be used to leverage unlabeled data for the target domain in domain generalization. The aim of Pseudo label DG for semantic segmentation is to enhance the quality of pseudo labels. This enables the model to be generalized well in unknown domains. Zhang et al. (2023) established *Multi-Target Domain Adaptation* (MTDA) framework which leverages implicit stylization and pseudo-labeling based on self-training to improve alignment between target domains. Kim et al. (2023b) presented WEDGE scheme to use the web-crawled images with their predicted pseudo labels for semantic segmentation. Yao et al. (2022) suggested a confident-aware cross pseudo supervision algorithm and Fourier transformed-based data augmentation to improve the quality of pseudo labels for unlabeled images from unknown distributions. Fourier transformation helps to obtain low-level static information and augment the image data using cross-domain information. Confidence-aware regularization helps to measure pseudo variances which can be used as a quality factor. Kundu et al. (2021) developed a conditional prior-enforcing auto-encoder to aid the client-side self-training. Hoyer et al. (2022) proposed a UDA-based method *DAFormer*. It comprises three ways where the quality of the pseudo-labels is improved by reducing the confirmation bias of self-training towards common classes through uncommon class sampling on the source domain.

5.11 Style transfer

Style transfer is a technique that is used to change image style while maintaining the content of the image. It is possible to build an overlap between source and target domains in the context of domain generalization (Su et al. 2022). Su et al. (2022) introduced a novel framework to perform an effective stylization with the preservation of fine-grained semantic clues for semantic segmentation. Wang et al. (2022b) proposed *Feature-based Style Randomization*(FSR) which helps to produce random styles to enhance the model robustness. Lee et al. (2022) proposed *feature stylization, content extension learning, style extension learning, semantic consistency regularization* by increasing both the content and style of the source domain to the wild. Zhao et al. (2022b) proposed (SHADE) based on two components *Style Consistency*(SC) and *Retrospection Consistency*(RC) to address domain shift. Gong et al. (2019) presented *domain flow generation*(DLOW) model, which is able in order to convert photos from the source domain into a random intermediate domain between the source and target domains. Fantauzzo et al. (2022) introduced *FedDrive*, a federated learning approach in semantic segmentation combined with style transfer techniques to improve their generalization.

6 Medical segmentation

In this section, popular applications for domain generalization (DG) in medical segmentation are discussed. Semantic Segmentation is widely used in medical imaging for precise diagnosis of diseases. Domain shift may pose a challenge for semantic segmentation. So there are ample applications of semantic segmentation based on domain generalization in the medical domain. Luo et al. (2023) proposed a single DG framework that is based on dual-level mixing for fundus image segmentation. Lyu et al. (2022) proposed an augmentation-based domain generalization method for renal image segmentation, this method generates novel domains for training. And novel proxy task maximizes the diversity between novel domains. Wang et al. (2020a) proposed a domain-oriented feature embedding to improve domain generalization for fundus image segmentation. Wang et al. (2019) also presented a method based on unsupervised domain adaptation via boundary-entropy-driven adversarial learning for optic disc (OD) and optic cup (OC) segmentation from fundus images. Liu et al. (2022) use T2-weighted MRIs from three public datasets including NCI-ISBI13 (Bloch et al. 2015), I2CVB (Lemaître et al. 2015), PROMISE (Litjens et al. 2014) for Prostate MRI segmentation and REFUGE (Orlando et al. 2020), DristhiGS (Sivaswamy et al. 2015), RIM-ONE-r3 (Fumero et al. 2011) for Fundus image segmentation. There are a few works on single-domain generalization (SGD). (Su et al. 2023) use cross-modality abdominal dataset (Landman et al. 2015) and cross-sequence cardiac dataset (Zhang et al. 2021b) for two single-source domain generalization (SDG) tasks. Yao et al. (2022) utilize M&M (Campello et al. 2021) and SCGM dataset (Prados et al. 2017) for multi-disease cardiac image segmentation. Xu et al. (2022a) conducted the experiment on cross-modality image segmentation with the abdominal CT scan (Landman et al. 2015) and MRI scans (Kavur et al. 2021). Bi et al. (2023) evaluate proposed MI-SegNet, a medical segmentation framework that is evaluated on ValS, TS1, TS2, TS3 (Říha et al. 2013) datasets. In Hu et al. (2021) Hu et al. demonstrate the effectiveness of the proposed *DAC*, *CAC* module on prostate segmentation using MRI, COVID-19 lesion segmentation using CT and OC/CD segmentation using color fundus image (Wang et al. 2020a; Liu et al. 2020; Tsai et al. 2021). Wang et al. (2020a) evaluate novel Domain-oriented Feature Embedding (DoFE) framework on optic cup (OC) / disc (OD) segmentation and vessel segmentation with retinal fundus image dataset. For semantic segmentation of hyperspectral images, Sellner et al. (2023) use 600 intraoperative Hyperspectral Images (HSI) under geometric domain shift. For left atrial (LA) segmentation, Li et al. (2021b) use late gadolinium-enhanced magnetic resonance imaging (LGE MRI) from MICCAI 2018 Atrial Segmentation Challenge (Pop et al. 2019) and ISBI 2012 Left Atrium Fibrosis and Scar Segmentation Challenge (Meng et al. 2020). Liu et al. (2021a) evaluate the proposed MixSearch framework on Composite, ISIC, CVC, Union, and CHAOS-CT datasets. Gu et al. (2021) demonstrate experimental results of proposed DCA-Net on multi-site prostate MRI segmentation using T2-weighted MRI dataset (Lemaître et al. 2015). Lyu et al. (2022) validate the proposed AADG framework on fundus vessel, OD/OC, retinal lesion, and OCTA vessel segmentation. Zhou et al. (2022b) demonstrate the effectiveness of the presented framework on Fundus (Wang et al. 2020a) and Prostate (Liu et al. 2020) segmentation task.

7 DGSS datasets and evaluation

7.1 Datasets

We describe most of the common and widely used benchmarks for *DGSS* task. *DGSS* benchmarks are divided into synthetic datasets and real-world datasets, there are some other rarely used datasets like ADE20k (Zhou et al. 2019), and MSeg (Lambert et al. 2020) available for *DGSS*, that are shown in Table 2.

GTA-V. *GTA-V* (Richter et al. 2016) is a synthetic semantic segmentation dataset that consists of nearly 25,000 densely labeled samples with 19 individual classes. The resolution of each sample is 1914×1052 pixels. It is extensively used in domain-generalized segmentation tasks.

Cityscapes. *Cityscapes* (Cordts et al. 2016) is a real-world driving dataset that consists of nearly 5000 labeled samples with 30 individual classes. The resolution of labeled samples is 2048×1024 pixels. In most of the *DG* literature, this dataset is used as a target set.

Mapillary. *Mapillary* (Neuhold et al. 2017) is a real-world semantic segmentation dataset that consists of 25000 labeled samples of 66 classes. The resolution of each sample is 1920×1024 pixels.

SYNTHIA. *SYNTHIA* (Ros et al. 2016) is a synthetic dataset as its name suggests. It is developed for semantic segmentation for urban scene understanding. It contains three different weather and illumination conditions, across three different road conditions (Highway, New York-like, and Old European Town). The majority of the work utilized 13 classes from this dataset, which has 9400 labeled samples. The resolution of each sample is 960×720 pixels.

KITTI. *KITTI* (Geiger et al. 2012) is a real-world semantic segmentation dataset that consists of nearly 400 labeled samples of 28 classes. The resolution of each sample is 1240×376 pixels.

IDD. *IDD* (Varma et al. 2019) is a real-world driving dataset that consists of nearly 10,000 labeled samples of 34 classes. The resolution of each sample is 1678×968 pixels.

BDD100k. *BDD100k* (Yu et al. 2020) is a real-world driving dataset that consists of 10000 labeled samples of 19 classes. The resolution of each sample is 1280×720 pixels.

ADCD. *ADCD* (Sakaridis et al. 2021) is a real-world driving dataset that consists of 4000 labeled samples of 19 classes. The resolution of each sample is 1920×1080 pixels (Table 1).

Table 1 Popular and extensively used datasets in domain generalization for semantic segmentation task

Dataset	Category	Classes	Annotated Samples	Resolution
<i>GTA-V</i> Richter et al. (2016)	Synthetic	19	24,966	1914×1052
<i>SYNTHIA</i> Ros et al. (2016)	Synthetic	13	200,000	960×720
<i>Cityscapes</i> Cordts et al. (2016)	Real	30	5000	2048×1024
<i>KITTI</i> Geiger et al. (2012)	Real	28	400	1240×376
<i>Mapillary</i> Neuhold et al. (2017)	Real	66	25,000	1920×1024
<i>IDD</i> Varma et al. (2019)	Real	34	10,004	1678×968
<i>BDD100k</i> Yu et al. (2020)	Real	19	10,000	1280×720
<i>ADCD</i> Sakaridis et al. (2021)	Real	19	4,000	1920×1080

Category denotes a real-world dataset or synthetic dataset

8 Future research directions

8.1 Variation in segmentation models

In most of the recent work we see, they used variants of the DeepLab model as a segmentation model. On the other hand, ResNet-50/ResNet-101 and VGG-16 are the backbone networks in most of the works. However, there is no work that utilizes the power of vision transformers in DG research. Utilizing the full power of vision transformers can lead to promising results in multiple challenging conditions. However, it is not well-answered how vision transformers can perform in domain gaps, hence these ViTs should be extensively explored as a backbone network.

8.2 Continual domain generalization

In many real-world applications, a system can encounter online data that belong to non-stationary distributions. So, to make a more robust segmentation model, generalization should be continuous against the non-stationary distributed data. It allows the model to learn and adapt data efficiently without catastrophic forgetting (Douillard et al. (2021)). Apparently, there is no work that has been done while focusing on this area.

8.3 Test-time generalization in segmentation

Most of the generalizations have been done in the training phase, we can also explore the inference phase to make it more concrete for real-world applications. It will allow us to leverage the full power of domain adaptation and generalization in a single framework. Test-time generalization helps to allow more flexibility and efficiency under limited resources (Wang et al. 2022a).

8.4 Large-scale benchmark

Most of the benchmarks are relatively small considering industrial applications. To achieve better generalization, we need a large-scale benchmark to overcome non-stationary shifts in the real-time target generalization. Currently, most of the segmentation tasks are done with the camera module, but autonomous vehicles or other related applications actively leverage the

8.5 Interpretability

Domain-invariant methods provide some interpretation in DG for segmentation tasks. However, other conventional DG methods do not provide enough or are not comprehensively interpretable. But in many cases, we need to understand how the comprehended

results are more close to the input space. This area can be explored for autonomous driving applications.

8.6 Vision-language models

Recently, vision-language models (VLMs) (Zhang et al. 2024a) have shown remarkable zero-shot transfer ability due to explicit vision-language pre-training (Gao et al. 2022; Bao et al. 2022) on multiple downstream tasks. Based on the applications, VLMs are becoming useful in OOD generalization tasks, hence multiple researches have explored plausible solutions for VLMs on domain generalization (Chen et al. 2024; Wang et al. 2024; Li et al. 2023b). However, most solutions focus general-purpose domain generalization, rather than specialized for the out-of-distribution segmentation tasks. So, this area can be explored due to the recent high potential of vision-language models in multiple applications.

8.7 Open vocabulary learning

Due to the recent emergence of vision-language models, open vocabulary learning (Wu et al. 2024) is proposed. Where models can discover categories beyond the training set category, it is no longer restricted to close-set classification. Recently, open vocabulary learning is adopted for domain adaptation tasks (Huang et al. 2023). However, it is widely considered for semantic segmentation tasks (Xu et al. 2023b; Liang et al. 2023; Xu et al. 2023a). So, this area also has a high-potential for DG-based semantic segmentation tasks. Future researches can be explored in this area.

8.8 Multimodal large-language models

Recently, multimodal large language models become the hotshot in AI research. It certainly has surprising power to deliver many downstream tasks, such as writing, generating codes or math reasoning. However, multimodal large language models are getting utilized in semantic segmentation tasks (Yang et al. 2024; Li et al. 2022a). Although, MLLMs are less popular in solving OOD problems (Zhang et al. 2024b). So, this area also can be explored particularly for DG for semantic segmentation.

9 Conclusion

In this paper, we comprehensively review the recent advances in domain generalization in semantic segmentation. In semantic segmentation tasks, domain adaptation is widely explored, but domain generalization is not well adopted. But generalization solves more tasks in more realistic and challenging scenarios. Our survey focuses on a very promising area of domain generalization in semantic segmentation. Most recent works have focused on domain adaptation in segmentation tasks, but the main challenge is large-scale deployment in industrial settings. We have explored recent generalization methods that are used in segmentation and provide a comprehensive overview of the whole scenario. We provide related background and methods that are extensively used in semantic segmentation alongside domain generalization. We also provide a critical analysis that is found in our observation as a future research direction in DG for segmentation. Based on the critical analysis, we recommend exploring variation in new baseline segmentation models, continual generalization in a real-world setting, test-time generalization, and interpretation. However, we believe that this survey will bring a new dimension to the community and interest in applying domain generalization in semantic segmentation tasks.

Appendix

See Tables 2, 3.

Table 2 Summary of state-of-art methods in DG for SS

Methods	References	Segmentation model	Backbone	Datasets
Data augmentation	Zhong et al. (2022)	DeepLabV3+	ResNet-101	G, C, B, M, S
Data augmentation	Kim et al. (2023a)	Minkoskinet	Minkoskinet-34	Semantic-KITTI, Waymo
Data augmentation	Shyam et al. (2022a)	DeepLabV2	ResNet-101	G, C, S
Data augmentation	Shen et al. (2023)	DeepLabV2	ResNet-101	G, C
Domain randomization	Gong et al. (2022)	DeepLabV2	ResNet-101	G, C, S
Domain randomization	Peng et al. (2021)	ResNet-101	MIT-B5	G, S, B, M
Domain randomization	Xiao et al. (2023)	Minkoskinet	Minkoskinet	Semantic-KITTI, Waymo
Domain randomization	Wu et al. (2022)	DeepLabV3+	ResNet-50	G, C, S, B, M, A
Domain randomization	Zhao et al. (2022b)	DeepLabV3+	ResNet-101	G, C, S, B, M
Adversarial Learning	Tjio et al. (2022)	DeepLabV2	ResNet-101	G, S, M
Adversarial Learning	Sankaranarayanan et al. (2018)	FCN-8 s	VGG-16	G, S, C
Self-Supervision	Vertens et al. (2020)	PSPNet	ResNet-50	Freiburg thermal, MF
Self-Supervision	Yang et al. (2023a)	DeepLabV3+	ResNet-50	G, S, C, B, M
Meta-Learning	Kim et al. (2022)	DeepLabV3+	ResNet-101	C, B, M, I, G, S
Meta-Learning	Zhang et al. (2020)	DeepLabV3+	ResNet-50	S, I, M, C, G
Meta-Learning	Gong et al. (2021)	DeepLabV2	VGG-16	G, S, B, C, KITTI, W
Meta-Learning	Shiau et al. (2021)	-	U-Net	C, G, S
Meta-Learning	Zhang et al. (2022a)	-	ResNet-50	C, G, S, I, M
Feature Disentanglement	Li et al. (2022b)	-	ResNet-50	C, K
Feature Disentanglement	Jin et al. (2021)	Faster R-CNN	ResNet-50	C, KITTI
Feature Normalization	Bahmani et al. (2021)	DeepLabV1	ResNet-101	C, B, I, S
Feature Normalization	Tjio et al. (2023)	DeepLabV2	ResNet-50	S, G, C
Domain Invariant	Xu et al. (2022b)	DeepLabV3+	ResNet-50	S, B, M, C, G
Domain Invariant	Liao et al. (2023)	DeepLabV2	VGG-16	G, C, M, B, S
Domain Invariant	He et al. (2023)	SegFormer	MIT-B5	C, G, S, B, A, M
Pseudo Label	Zhang et al. (2023)	DeepLabV2	ResNet-101	G, C, M, I
Pseudo Label	Kim et al. (2023b)	DeepLabV2	ResNet-101	G, C, B, M, S

Table 2 (continued)

Methods	References	Segmentation model	Backbone	Datasets
Pseudo Label	Kundu et al. (2021)	DeepLabV2	ResNet-101	G, C, S
Pseudo Label	Hoyer et al. (2022)	MMSegmentation	MiT-B5	C, G, S
Style transfer	Su et al. (2022)	DeepLabV3+	ResNet-50	C, B, M, G, S
Style transfer	Lee et al. (2022)	DeepLabV3+	ResNet-101	C, B, M, G, S
Style transfer	Zhao et al. (2022b)	DeepLabV3+	ResNet-101	G, S, C, B, M
Style transfer	Gong et al. (2019)	AdaptSegNet	ResNet-101	C, G, KITTI, W, B
Style transfer	Fantauzzo et al. (2022)	BiSeNetV2	–	C, IDDA

G, S, B, M, C, I, and W as GTA-V, Synthia, BDD, Mapillary, Cityscape, IDD, and Waymo respectively

Table 3 Summary of state-of-art performances in mIoU in DG for SS

Methods	References	GTA-V	Synthia	BDD	Mapillary	Cityscape	IDD	Waymo	S-KITTI
Data augmentation	Zhong et al. (2022)	Source	-	41.71	46.69	45.62	-	-	-
Data augmentation	Kim et al. (2023a)	Source	-	-	-	-	-	Source	51.23
Data augmentation	Shyam et al. (2022a)	Source	-	-	-	57.50	-	-	-
Data augmentation	Shen et al. (2023)	Source	-	-	-	62.70	-	-	-
Domain randomization	Gong et al. (2022)	Source	-	-	-	55.37	-	-	-
Domain randomization	Peng et al. (2021)	Source	-	39.60	39.10	43.70	-	-	-
Domain randomization	Wu et al. (2022)	Source	28.34	37.54	40.64	42.96	-	-	-
Domain Generation	Zhao et al. (2022b)	Source	-	43.66	45.50	46.66	-	-	-
Adversarial Learning	Tjio et al. (2022)	Source	-	-	44.44	41.30	-	-	-
Adversarial Learning	Sankaranarayanan et al. (2018)	Source	-	-	-	37.10	-	-	-
Self-Supervision	Yang et al. (2023a)	Source	-	41.14	48.89	46.34	-	-	-
Meta-Learning	Kim et al. (2022)	Source	Source	38.07	42.70	44.51	-	-	-
Meta-Learning	Zhang et al. (2020)	Source	-	-	-	44.20	-	-	-
Meta-Learning	Gong et al. (2021)	Source	-	-	29.40	-	-	-	-
Feature Disentanglement	Jin et al. (2021)	Source	-	-	-	42.68	-	-	-
Feature Normalization	Bahmani et al. (2021)	Source	-	40.21	47.39	45.62	-	-	-
Feature Normalization	Tjio et al. (2023)	Source	-	-	-	43.13	-	-	-
Domain Invariant	Xu et al. (2022b)	Source	-	44.10	51.10	45.30	-	-	-
Domain Invariant	Liao et al. (2023)	Source	32.10	38.60	42.80	42.60	-	-	-
Domain Invariant	He et al. (2023)	Source	-	39.15	41.60	41.04	-	-	-
Pseudo Label	Zhang et al. (2023)	Source	-	-	-	47.10	49.30	-	-
Pseudo Label	Kim et al. (2023b)	Source	-	41.10	48.10	45.20	-	-	-
Pseudo Label	Kundh et al. (2021)	Source	-	-	-	53.40	-	-	-
Pseudo Label	Hoyer et al. (2022)	Source	-	-	-	45.60	-	-	-
Style transfer	Su et al. (2022)	Source	30.68	40.52	43.53	40.84	-	-	-
Style transfer	Lee et al. (2022)	Source	32.51	41.73	47.08	45.79	-	-	-
Style transfer	Zhao et al. (2022b)	Source	27.25	34.49	40.75	39.18	-	-	-

Table 3 (continued)

Methods	References	GTA-V	Synthia	BDD	Mapillary	Cityscape	IDD	Waymo	S-KITTI
Style transfer	Gong et al. (2019)	Source	-	-	-	42.30	-	-	-

Here, we mentioned the best results for the most commonly used datasets like GTA-V, Synthia, BDD, Mapillary, Cityscape, IDD, Waymo and Semantic-KITTI (S-KITTI) respectively

Acknowledgements This work was supported by the Hallym University Research Fund, 2022, under Grant HRF-202210-001.

Author contributions T.H.R., R.M., and E.G. wrote the whole manuscript. Y.W. Ko has supervised and reviewed the whole writing. Y.W. Ko and J.G.L. managed the funding for this paper.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aakerberg A, Johansen AS, Nasrollahi K, Moeslund TB (2021) Single-loss multi-task learning for improving semantic segmentation using super-resolution. In: Computer analysis of images and patterns: 19th International Conference, CAIP 2021, Virtual Event, 28–30 September 2021, Proceedings, Part II 19. Springer, Cham, pp 403–411
- Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(12):2481–2495
- Bae W, Noh J, Asadabadi MJ, Sutherland DJ (2022) One weird trick to improve your semi-weakly supervised semantic segmentation model. arXiv preprint. [arXiv:2205.01233](https://arxiv.org/abs/2205.01233)
- Bahmani S, Hahn O, Zamfir ES, Araslanov N, Roth S (2021) Adaptive generalization for semantic segmentation. arXiv preprint. [arXiv:2208.05788](https://arxiv.org/abs/2208.05788)
- Bao H, Wang W, Dong L, Liu Q, Mohammed OK, Aggarwal K, Som S, Piao S, Wei F (2022) Vlmo: unified vision-language pre-training with mixture-of-modality-experts. *Adv Neural Inform Proc Syst* 35:32897–32912
- Ben-David S, Blitzer J, Crammer K, Kulesza A, Pereira F, Vaughan JW (2010) A theory of learning from different domains. *Mach Learn* 79:151–175
- Bi Y, Jiang Z, Clarenbach R, Ghotbi R, Karlas A, Navab N (2023) MI-SegNet: mutual information-based us segmentation for unseen domain generalization. arXiv preprint. [arXiv:2303.12649](https://arxiv.org/abs/2303.12649)
- Bischke B, Helber P, Folz J, Borth D, Dengel A (2019) Multi-task learning for segmentation of building footprints with deep neural networks. In: 2019 IEEE international conference on image processing (ICIP). IEEE, pp 1480–1484
- Blanchard G, Deshmukh AA, Dogan Ü, Lee G, Scott C (2021) Domain generalization by marginal transfer learning. *J Mach Learn Res* 22(1):46–100
- Bloch N, Madabhushi A, Huisman H, Freymann J, Kirby J, Grauer M, Enquobahrie A, Jaffe C, Clarke L, Farahani K (2015) NCI-ISBI 2013 challenge: automated segmentation of prostate structures. *Cancer Imaging Archive*. <https://doi.org/10.7937/K9/TCIA.2015.zF0vIOPv>
- Campello VM, Gkontra P, Izquierdo C, Martin-Isla C, Sojoudi A, Full PM, Maier-Hein K, Zhang Y, He Z, Ma J (2021) Multi-centre, multi-vendor and multi-disease cardiac segmentation: the M&Ms challenge. *IEEE Trans Med Imaging* 40(12):3543–3554
- Chen Z, Fu Y, Chen K, Jiang Y-G (2019) Image block augmentation for one-shot learning. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, pp 3379–3386
- Chen J, Gao B-B, Lu Z, Xue J-H, Wang C, Liao Q (2021a) Apanet: adaptive prototypes alignment network for few-shot semantic segmentation. arXiv preprint. [arXiv:2111.12263](https://arxiv.org/abs/2111.12263)
- Chen Y, Ouyang X, Zhu K, Agam G (2021b) Semi-supervised domain adaptation for semantic segmentation. arXiv preprint [arXiv:2110.10639](https://arxiv.org/abs/2110.10639)
- Chen T, Yao Y, Tang J (2023a) Multi-granularity denoising and bidirectional alignment for weakly supervised semantic segmentation. *IEEE Trans Image Proc*. <https://doi.org/10.1109/TIP.2023.3275913>

- Chen J, Lu J, Zhu X, Zhang L (2023b) Generative semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 7111–7120
- Chen Z, Wang W, Zhao Z, Su F, Men A, Meng H (2024) PracticalDG: perturbation distillation on vision-language models for hybrid domain generalization. arXiv preprint. [arXiv:2404.09011](https://arxiv.org/abs/2404.09011)
- Cheng B, Collins MD, Zhu Y, Liu T, Huang TS, Adam H, Chen L-C (2020) Panoptic-DeepLab: a simple, strong, and fast baseline for bottom-up panoptic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 12475–12485
- Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S, Schiele B (2016) The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3213–3223
- Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, pp 248–255
- Ding H, Sun C, Tang H, Cai D, Yan Y (2023) Few-shot medical image segmentation with cycle-resemblance attention. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 2488–2497
- Dou Q, Castro D, Kamnitsas K, Glocker B (2019) Domain generalization via model-agnostic learning of semantic features. *Adv Neural Inform Proc Syst*. <https://doi.org/10.48550/arXiv.1910.13580>
- Douillard A, Chen Y, Dapogny A, Cord M (2021) Plop: Learning without forgetting for continual semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4040–4050
- Dwibedi D, Misra I, Hebert M (2017) Cut, paste and learn: surprisingly easy synthesis for instance detection. In: Proceedings of the IEEE international conference on computer vision, pp 1301–1310
- Fantauzzo L, Fani E, Calderola D, Tavera A, Cermelli F, Ciccone M, Caputo B (2022) FedDrive: generalizing federated learning to semantic segmentation in autonomous driving. In: 2022 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 11504–11511
- Fumero F, Alayón S, Sanchez JL, Sigut J, Gonzalez-Hernandez M (2011) RIM-ONE: an open retinal image database for optic nerve evaluation. In: 2011 24th International symposium on computer-based medical systems (CBMS). IEEE, pp 1–6
- Gal Y, Ghahramani Z (2015) Bayesian convolutional neural networks with bernoulli approximate variational inference. arXiv preprint. [arXiv:1506.02158](https://arxiv.org/abs/1506.02158)
- Ganin Y, Lempitsky V (2015) Unsupervised domain adaptation by backpropagation. In: International conference on machine learning. PMLR, pp 1180–1189
- Gao Y, Liu J, Xu Z, Zhang J, Li K, Ji R, Shen C (2022) PyramidCLIP: hierarchical feature alignment for vision-language model pretraining. *Adv Neural Inform Proc Syst* 35:35959–35970
- Geiger A, Lenz P, Urtasun R (2012) Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, pp 3354–3361
- Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 580–587
- Gong R, Li W, Chen Y, Gool LV (2019) Dlow: domain flow for adaptation and generalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2477–2486
- Gong R, Chen Y, Paudel DP, Li Y, Chhatkuli A, Li W, Dai D, Van Gool L (2021) Cluster, split, fuse, and update: meta-learning for open compound domain adaptive semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 8344–8354
- Gong R, Wang Q, Dai D, Van Gool L (2022) One-shot domain adaptive and generalizable semantic segmentation with class-aware cross-domain transformers. arXiv preprint. [arXiv:2212.07292](https://arxiv.org/abs/2212.07292)
- Graham S, Vu QD, Jahanifar M, Raza SEA, Minhas F, Snead D, Rajpoot N (2023) One model is all you need: multi-task learning enables simultaneous histology image segmentation and classification. *Med Image Anal* 83:102685
- Gu R, Zhang J, Huang R, Lei W, Wang G, Zhang S (2021) Domain composition and attention for unseen-domain generalizable medical image segmentation. In: Medical image computing and computer assisted intervention—MICCAI 2021: 24th international conference, Strasbourg, France, 27 September–1 October 2021, proceedings, Part III 24. Springer, Berlin, pp 241–250
- Guo G, Zhang N (2019) A survey on deep learning based face recognition. *Comput Vis Image Underst* 189:102805
- Guo Y, Liu Y, Georgiou T, Lew MS (2018) A review of semantic segmentation using deep neural networks. *Int J Multimedia Inform Retr* 7:87–93
- Ham S, Kim M, Lee S, Wang C-B, Ko B, Kim N (2023) Improvement of semantic segmentation through transfer learning of multi-class regions with convolutional neural networks on supine and prone breast mri images. *Sci Rep* 13(1):6877

- He J, Jia X, Chen S, Liu J (2021) Multi-source domain adaptation with collaborative learning for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11008–11017
- He P, Jiao L, Shang R, Liu X, Liu F, Yang S, Zhang X, Wang S (2023) A patch diversity transformer for domain generalized semantic segmentation. *IEEE Trans Neural Netw Learn Syst.* <https://doi.org/10.1109/TNNLS.2023.3274760>
- Honarvar Nazari N, Kovashka A (2020) Domain generalization using shape representation. In: Computer vision—ECCV 2020 workshops: Glasgow, UK, 23–28 August 2020, proceedings, Part I 16, 666–670. Springer, Berlin
- Hoyer L, Dai D, Van Gool, L (2022) DAFormer: improving network architectures and training strategies for domain-adaptive semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 9924–9935
- Hoyer L, Dai D, Wang Q, Chen Y, Van Gool L (2023) Improving semi-supervised and domain-adaptive semantic segmentation with self-supervised depth estimation. *Int J Comput Vis* 131(8):2070–2096
- Hu S, Liao Z, Zhang J, Xia Y (2021) Domain and content adaptive convolution based multi-source domain generalization for medical image segmentation. arXiv preprint. [arXiv:2109.05676](https://arxiv.org/abs/2109.05676)
- Huang J, Zhang J, Qiu H, Jin S, Lu S (2023) Prompt ensemble self-training for open-vocabulary domain adaptation. arXiv preprint. [arXiv:2306.16658](https://arxiv.org/abs/2306.16658)
- Jin X, Lan C, Zeng W, Chen Z (2021) Style normalization and restitution for domain generalization and adaptation. *IEEE Trans Multimedia* 24:3636–3651
- Kalluri T, Chandraker M (2022) Cluster-to-adapt: Few shot domain adaptation for semantic segmentation across disjoint labels. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 4121–4131
- Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T (2020) Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 8110–8119
- Kavur AE, Gezer NS, Barış M, Aslan S, Conze P-H, Groza V, Pham DD, Chatterjee S, Ernst P, Özkan S (2021) Chaos challenge-combined (CT-MR) healthy abdominal organ segmentation. *Med Image Anal* 69:101950
- Kendall A, Badrinarayanan V, Cipolla R (2015) Bayesian SegNet: model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. arXiv preprint. [arXiv:1511.02680](https://arxiv.org/abs/1511.02680)
- Khosla C, Saini BS (2020) Enhancing performance of deep learning models with different data augmentation techniques: a survey. In: 2020 International conference on intelligent engineering and management (ICIEM). IEEE, pp 79–85
- Kim J, Lee J, Park J, Min D, Sohn K (2022) Pin the memory: Learning to generalize semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4350–4360
- Kim H, Kang Y, Oh C, Yoon K-J (2023a) Single domain generalization for lidar semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 17587–17598
- Kim N, Son T, Park J, Lan C, Zeng W, Kwak S (2023b) Wedge: web-image assisted domain generalization for semantic segmentation. In: 2023 IEEE international conference on robotics and automation (ICRA). IEEE, pp 9281–9288
- Kirillov A, Wu Y, He K, Girshick R (2020) PointRend: Image segmentation as rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9799–9808
- Kumar Singh K, Jae Lee Y (2017) Hide-and-seek: forcing a network to be meticulous for weakly-supervised object and action localization. In: Proceedings of the IEEE international conference on computer vision, pp 3524–3533
- Kundu JN, Kulkarni A, Singh A, Jampani V, Babu RV (2021) Generalize then adapt: source-free domain adaptive semantic segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 7046–7056
- Kweon H, Yoon S-H, Yoon K-J (2023) Weakly supervised semantic segmentation via adversarial learning of classifier and reconstructor. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 11329–11339
- Lambert J, Liu Z, Sener O, Hays J, Koltun V (2020) MSEG: a composite dataset for multi-domain semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2879–2888
- Landman B, Xu Z, Igelsias J, Styner M, Langerak T, Klein A (2015) MICCAI multi-atlas labeling beyond the cranial vault-workshop and challenge. In: Proc. MICCAI multi-atlas labeling beyond cranial vault-workshop challenge, vol 5, p 12

- Lee, S., Hyun, J., Seong, H., Kim, E (2021) Unsupervised domain adaptation for semantic segmentation by content transfer. In: Proceedings of the AAAI conference on artificial intelligence, vol 35, pp 8306–8315
- Lee S, Seong H, Lee S, Kim E (2022) Wildnet: Learning domain generalized semantic segmentation from the wild. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 9936–9946
- Lei S, Zhang X, He J, Chen F, Du B, Lu C-T (2022) Cross-domain few-shot semantic segmentation. In: European conference on computer vision. Springer, Cham, pp 73–90
- Lemaître G, Martí R, Freixenet J, Vilanova JC, Walker PM, Meriaudeau F (2015) Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. *Comput Biol Med* 60:8–31
- Liang F, Wu B, Dai X, Li K, Zhao Y, Zhang H, Zhang P, Vajda P, Marculescu D (2023) Open-vocabulary semantic segmentation with mask-adapted clip. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 7061–7070
- Liao M, Tian S, Zhang Y, Hua G, Zou W, Li X (2023) Domain-invariant information aggregation for domain generalization semantic segmentation. *Neurocomputing* 546:126273
- Li X, Li X, Zhang L, Cheng G, Shi J, Lin Z, Tan S, Tong Y (2020a) Improving semantic segmentation via decoupled body and edge supervision. In: Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII 16,435–452. Springer
- Li X, You A, Zhu Z, Zhao H, Yang M, Yang K, Tan S, Tong Y (2020b) Semantic flow for fast and accurate scene parsing. In: Computer vision—ECCV 2020: 16th European conference, Glasgow, UK, 23–28 August 2020, proceedings, Part I 16. Springer, pp 775–793
- Li D, Yang J, Kreis K, Torralba A, Fidler S (2021a) Semantic segmentation with generative models: Semi-supervised learning and strong out-of-domain generalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 8300–8311
- Li L, Zimmer VA, Schnabel JA, Zhuang X (2021b) Atrialgeneral: domain generalization for left atrial segmentation of multi-center LGE MRIS. In: Medical image computing and computer assisted intervention—MICCAI 2021: 24th international conference, Strasbourg, France, 27 September–1 October 2021, proceedings, Part VI 24. Springer, Cham, pp 557–566
- Li Y, Data GWP, Fu Y, Hu Y, Priscarari VA (2021c) Few-shot semantic segmentation with self-supervision from pseudo-classes. arXiv preprint. [arXiv:2110.11742](https://arxiv.org/abs/2110.11742)
- Li B, Weinberger KQ, Belongie S, Koltun V, Ranftl R (2022a) Language-driven semantic segmentation. arXiv preprint. [arXiv:2201.03546](https://arxiv.org/abs/2201.03546)
- Li X, Li L, Guo X (2022b) Synthetic-to-real generalization for semantic segmentation. In: 2022 IEEE international conference on multimedia and expo (ICME). IEEE, pp 1–6
- Li X, Ding H, Zhang W, Yuan H, Pang J, Cheng G, Chen K, Liu Z, Loy CC (2023a) Transformer-based visual segmentation: a survey. arXiv preprint. [arXiv:2304.09854](https://arxiv.org/abs/2304.09854)
- Li J, Gao M, Wei L, Tang S, Zhang W, Li M, Ji W, Tian Q, Chua T-S, Zhuang Y (2023b) Gradient-regulated meta-prompt learning for generalizable vision-language models. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 2551–2562
- Litjens G, Toth R, Van De Ven W, Hoeks C, Kerkstra S, Van Ginneken B, Vincent G, Guillard G, Birbeck N, Zhang J (2014) Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. *Med Image Anal* 18(2):359–373
- Liu Q, Dou Q, Heng P-A (2020) Shape-aware meta-learning for generalizing prostate mri segmentation to unseen domains. In: Medical image computing and computer assisted intervention—MICCAI 2020: 23rd international conference, Lima, Peru, 4–8 October 2020, proceedings, Part II 23. Springer, pp 475–485
- Liu L, Wen Z, Liu S, Zhou H-Y, Zhu H, Xie W, Shen L, Ma K, Zheng Y (2021a) Mixsearch: Searching for domain generalized medical image segmentation architectures. arXiv preprint. [arXiv:2102.13280](https://arxiv.org/abs/2102.13280)
- Liu Y, Zhang W, Wang J (2021b) Source-free domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 1215–1224
- Liu Q, Chen C, Dou Q, Heng P-A (2022) Single-domain generalization in medical image segmentation via test-time adaptation from shape dictionary. In: Proceedings of the AAAI Conference on Artificial Intelligence, 36:1756–1764
- Liu Y-P, Zeng D, Li Z, Chen P, Liang R (2023) SS-Norm: spectral-spatial normalization for single-domain generalization with application to retinal vessel segmentation. *IET Image Proc* 17(7):2168–2181
- Lu Y, Sarkis M, Lu G (2020) Multi-task learning for single image depth estimation and segmentation based on unsupervised network. In: 2020 IEEE international conference on robotics and automation (ICRA). IEEE, pp 10788–10794

- Luo X, Chen W, Li C, Zhou B, Tan Y (2023) Domain generalized fundus image segmentation via dual-level mixing. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1–5. IEEE
- Lyu J, Zhang Y, Huang Y, Lin L, Cheng P, Tang X (2022) Aadg: automatic augmentation for domain generalization on retinal image segmentation. *IEEE Trans Med Imaging* 41(12):3699–3711
- Meng Q, Matthew J, Zimmer VA, Gomez A, Lloyd DF, Rueckert D, Kainz B (2020) Mutual information-based disentangled neural networks for classifying unseen categories in different domains: application to fetal ultrasound imaging. *IEEE Trans Med Imaging* 40(2):722–734
- Misra I, Shrivastava A, Gupta A, Hebert M (2016) Cross-stitch networks for multi-task learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3994–4003
- Moreno-Torres JG, Raeder T, Alaiz-Rodríguez R, Chawla NV, Herrera F (2012) A unifying view on dataset shift in classification. *Pattern Recognit* 45(1):521–530
- Muandet K, Baldazzi D, Schölkopf B (2013) Domain generalization via invariant feature representation. In: International conference on machine learning. PMLR, pp 10–18
- Neuhold G, Ollmann T, Rota Bulò S, Kotschieder P (2017) The mapillary vistas dataset for semantic understanding of street scenes. In: Proceedings of the IEEE international conference on computer vision, pp 4990–4999
- Nigam I, Huang C, Ramanan D (2018) Ensemble knowledge transfer for semantic segmentation. In: 2018 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 1499–1508
- Noh H, Hong S, Han B (2015) Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE international conference on computer vision, pp 1520–1528
- Novosel J, Viswanath P, Arsenali B (2019) Boosting semantic segmentation with multi-task self-supervised learning for autonomous driving applications. In: Proceedings of NeurIPS—workshops, vol 3
- Orlando JI, Fu H, Breda JB, Van Keer K, Bathula DR, Diaz-Pinto A, Fang R, Heng P-A, Kim J, Lee J (2020) Refuge challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med Image Anal* 59:101570
- Pambala AK, Dutta T, Biswas S (2021) Sml: semantic meta-learning for few-shot semantic segmentation. *Pattern Recognit Lett* 147:93–99
- Pan SJ, Yang Q (2009) A survey on transfer learning. *IEEE Trans Knowl Data Eng* 22(10):1345–1359
- Peng D, Lei Y, Liu L, Zhang P, Liu J (2021) Global and local texture randomization for synthetic-to-real semantic segmentation. *IEEE Trans Image Proc* 30:6594–6608
- Pop M, Sermesant M, Zhao J, Li S, McLeod K, Young A, Rhode K, Mansi T (2019) Statistical atlases and computational models of the heart. In: Atrial segmentation and LV quantification challenges: 9th international workshop, STACOM 2018, held in conjunction with MICCAI 2018, Granada, Spain, 16 September 2018, revised selected papers, vol 11395. Springer, Cham
- Prados F, Ashburner J, Blaiotta C, Brosch T, Carballido-Gamio J, Cardoso MJ, Conrad BN, Datta E, Dávid G, De Leener B (2017) Spinal cord grey matter segmentation challenge. *Neuroimage* 152:312–329
- Qiao F, Zhao L, Peng X (2020) Learning to learn single domain generalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, PP 12556–12565
- Rao Q, Frtunik J (2018) Deep learning for self-driving cars: chances and challenges. In: Proceedings of the 1st international workshop on software engineering for AI in autonomous systems, PP 35–38
- Recht B, Roelofs R, Schmidt L, Shankar V (2019) Do ImageNet classifiers generalize to imagenet? In: International conference on machine learning. PMLR, pp 5389–5400
- Richter SR, Vineet V, Roth S, Koltun V (2016) Playing for data: Ground truth from computer games. In: Computer vision—ECCV 2016: 14th European conference, Amsterdam, The Netherlands, 11–14 October 2016, proceedings, Part II 14. Springer, Cham, pp 102–118
- Říha K, Mašek J, Burget R, Beneš R, Závodná E (2013) Novel method for localization of common carotid artery transverse section in ultrasound images using modified Viola–Jones detector. *Ultrasound Med Biol* 39(10):1887–1902
- Rong S, Tu B, Wang Z, Li J (2023) Boundary-enhanced co-training for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 19574–19584
- Ros G, Sellart L, Materzynska J, Vazquez D, Lopez AM (2016) The synthia dataset: a large collection of synthetic images for semantic segmentation of urban scenes. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3234–3243
- Roth HR, Shen C, Oda H, Oda M, Hayashi Y, Misawa K, Mori K (2018) Deep learning and its application to medical image segmentation. *Med Imaging Technol* 36(2):63–71
- Sakaridis C, Dai D, Van Gool L (2021) ACDC: the adverse conditions dataset with correspondences for semantic driving scene understanding. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 10765–10775

- Sankaranarayanan S, Balaji Y, Jai, A, Lim SN, Chellappa R (2017) Unsupervised domain adaptation for semantic segmentation with GANS. arXiv preprint. [arXiv:1711.06969v1](https://arxiv.org/abs/1711.06969v1)
- Sankaranarayanan S, Balaji Y, Jain A, Lim SN, Chellappa R (2018) Learning from synthetic data: addressing domain shift for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3752–3761
- Sellner J, Seidlitz S, Studier-Fischer A, Motta A, Özdemir B, Müller-Stich BP, Nickel F, Maier-Hein L (2023) Semantic segmentation of surgical hyperspectral images under geometric domain shifts. arXiv preprint. [arXiv:2303.10972](https://arxiv.org/abs/2303.10972)
- Shen F, Gurram A, Liu Z, Wang H, Knoll A (2023) Diga: Distil to generalize and then adapt for domain adaptive semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 15866–15877
- Shiau Z-Y, Lin W-W, Lin C-S, Wang Y-CF (2021) Meta-learned feature critics for domain generalized semantic segmentation. In: 2021 IEEE international conference on image processing (ICIP). IEEE, pp 2244–2248
- Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. *J Big Data* 6(1):1–48
- Shyam P, Bangunharcana A, Yoon K-J, Kim K-S (2022a) DGSS: domain generalized semantic segmentation using iterative style mining and latent representation alignment. arXiv preprint. [arXiv:2202.13144](https://arxiv.org/abs/2202.13144)
- Shyam P, Yoon K-J, Kim K-S (2022b) Multi-source domain alignment for domain invariant segmentation in unknown targets. In: 2022 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 309–316
- Sivaswamy J, Krishnadas S, Chakravarty A, Joshi G, Tabish AS (2015) A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomed Imaging Data Pap* 2(1):1004
- Souly N, Spampinato C, Shah M (2017) Semi and weakly supervised semantic segmentation using generative adversarial network. arXiv preprint. [arXiv:1703.09695](https://arxiv.org/abs/1703.09695)
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15(1):1929–1958
- Strudel R, Garcia R, Laptev I, Schmid C (2021) Segmenter: transformer for semantic segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 7262–7272
- Su S, Wang H, Yang M (2022) Consistency learning based on class-aware style variation for domain generalizable semantic segmentation. In: Proceedings of the 30th ACM International conference on multimedia, pp 6029–6038
- Su Z, Yao K, Yang X, Huang K, Wang Q, Sun J (2023) Rethinking data augmentation for single-source domain generalization in medical image segmentation. In: Proceedings of the AAAI conference on artificial intelligence, vol 37, pp 2366–2374
- Sun R, Zhu X, Wu C, Huang C, Shi J, Ma L (2019) Not all areas are equal: transfer learning for semantic segmentation via hierarchical region selection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4360–4369
- Taori R, Dave A, Shankar V, Carlini N, Recht B, Schmidt L (2020) Measuring robustness to natural distribution shifts in image classification. *Adv Neural Inform Proc Syst* 33:18583–18599
- Tian P, Wu Z, Qi L, Wang L, Shi Y, Gao Y (2020) Differentiable meta-learning model for few-shot semantic segmentation. In: Proceedings of the AAAI Conference on artificial intelligence, vol 34, pp 12087–12094
- Tjio G, Liu P, Zhou JT, Goh RSM (2022) Adversarial semantic hallucination for domain generalized semantic segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 318–327
- Tjio G, Liu P, Kwok C-K, Tianyi Zhou J (2023) Adaptive stylization modulation for domain generalized semantic segmentation. arXiv e-prints, 2304
- Toldo M, Michieli U, Zanuttigh P (2022) Learning with style: continual semantic segmentation across tasks and domains. arXiv preprint. [arXiv:2210.07016](https://arxiv.org/abs/2210.07016)
- Tsai EB, Simpson S, Lungren MP, Hershman M, Roshkovan L, Colak E, Erickson BJ, Shih G, Stein A, Kalpathy-Cramer J (2021) The RSNA international covid-19 open radiology database (RICORD). *Radiology* 299(1):204–213
- Van Engelen JE, Hoos HH (2020) A survey on semi-supervised learning. *Mach Learn* 109(2):373–440
- Varma G, Subramanian A, Nambodiri A, Chandraker M, Jawahar C (2019) IDD: a dataset for exploring problems of autonomous navigation in unconstrained environments. In: 2019 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 1743–1751

- Vertens J, Zörn J, Burgard W (2020) HeatNet: bridging the day-night domain gap in semantic segmentation with thermal images. In: 2020 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 8461–8468
- Wang S, Yu L, Li K, Yang X, Fu C-W, Heng P-A (2019) Boundary and entropy-driven adversarial learning for fundus image segmentation. In: Medical image computing and computer assisted intervention—MICCAI 2019: 22nd international conference, Shenzhen, China, 13–17 October 2019, proceedings, Part I 22. Springer, Cham, pp 102–110
- Wang S, Yu L, Li K, Yang X, Fu C-W, Heng P-A (2020a) DOFE: domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets. *IEEE Trans Med Imaging* 39(12):4237–4248
- Wang Z, Wei Y, Feris R, Xiong J, Hwu W-M, Huang TS, Shi H (2020b) Alleviating semantic-level shift: a semi-supervised domain adaptation method for semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp 936–937
- Wang H, Zhu Y, Adam H, Yuille A, Chen L-C (2021) MaX-DeepLab: end-to-end panoptic segmentation with mask transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5463–5474
- Wang J, Lan C, Liu C, Ouyang Y, Qin T, Lu W, Chen Y, Zeng W, Yu P (2022a) Generalizing to unseen domains: a survey on domain generalization. *IEEE Trans Knowl Data Eng* 8:8052–8072
- Wang Y, Qi L, Shi Y, Gao Y (2022b) Feature-based style randomization for domain generalization. *IEEE Trans Circuits Syst Video Technol* 32(8):5495–5509
- Wang H, Shen Y, Fei J, Li W, Wu L, Wang Y, Zhang Z (2023) Pulling target to source: a new perspective on domain adaptive semantic segmentation. *arXiv preprint. arXiv:2305.13752*
- Wang L, Jin Y, Chen Z, Wu J, Li M, Lu Y, Wang H (2024) Transitive vision-language prompt learning for domain generalization. *arXiv preprint. arXiv:2404.18758*
- Wu Z, Wu X, Zhang X, Ju L, Wang S (2022) SiamDoGe: domain generalizable semantic segmentation using Siamese network. In: European conference on computer vision. Springer, Cham, pp 603–620
- Wu J, Li X, Xu S, Yuan H, Ding H, Yang Y, Li X, Zhang J, Tong Y, Jiang X et al (2024) Towards open vocabulary learning: a survey. *IEEE Trans Pattern Anal Mach Intell.* <https://doi.org/10.48550/arXiv.2306.15880>
- Xiao A, Huang J, Xuan W, Ren R, Liu K, Guan D, El Saddik A, Lu S, Xing EP (2023) 3D semantic segmentation in the wild: Learning generalized models for adverse-condition point clouds. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 9382–9392
- Xie E, Wang W, Yu Z, Anandkumar A, Alvarez JM, Luo P (2021) Segformer: simple and efficient design for semantic segmentation with transformers. *Adv Neural Inform Proc Syst* 34:12077–12090
- Xie B, Li S, Li M, Liu CH, Huang G, Wang G (2023) SEPICO: semantic-guided pixel contrast for domain adaptive semantic segmentation. *IEEE Trans Pattern Anal Mach Intell.* <https://doi.org/10.1109/TPAMI.2023.3237740>
- Xu Y, Du B, Zhang L, Zhang Q, Wang G, Zhang L (2019) Self-ensembling attention networks: addressing domain shift for semantic segmentation. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, pp 5581–5588
- Xu, Q., Zhang, R., Zhang, Y, Wang Y, Tian Q (2021) A Fourier-based framework for domain generalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 14383–14392
- Xu Y, Xie S, Reynolds M, Ragoza M, Gong M, Batmanghelich K (2022a) Adversarial consistency for single domain generalization in medical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 671–681
- Xu Q, Yao L, Jiang Z, Jiang G, Chu W, Han W, Zhang W, Wang C, Tai Y (2022b) DURL: domain-invariant representation learning for generalizable semantic segmentation. In: Proceedings of the AAAI conference on artificial intelligence, vol 36, pp 2884–2892
- Xu J, Hou J, Zhang Y, Feng R, Wang Y, Qiao Y, Xie W (2023a) Learning open-vocabulary semantic segmentation models from natural language supervision. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2935–2944
- Xu M, Zhang Z, Wei F, Hu H, Bai X (2023b) Side adapter network for open-vocabulary semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2945–2954
- Yang W, Zhang J, Chen Z, Xu Z (2021) An efficient semantic segmentation method based on transfer learning from object detection. *IET Image Proc* 15(1):57–64
- Yang C-Y, Kuo Y-J, Hsu C-T (2022a) Source free domain adaptation for semantic segmentation via distribution transfer and adaptive class-balanced self-training. In: 2022 IEEE international conference on multimedia and expo (ICME). IEEE, pp 1–6

- Yang S, Xiao W, Zhang M, Guo S, Zhao J, Shen F (2022b) Image data augmentation for deep learning: A survey. arXiv preprint [arXiv:2204.08610](https://arxiv.org/abs/2204.08610)
- Yang Z, Yu H, Sun W, Mian A et al (2022c) Domain-invariant prototypes for semantic segmentation. arXiv preprint [arXiv:2208.06087](https://arxiv.org/abs/2208.06087)
- Yang L, Gu X, Sun J (2023a) Generalized semantic segmentation by self-supervised source domain projection and multi-level contrastive learning. arXiv preprint. [arXiv:2303.01906](https://arxiv.org/abs/2303.01906)
- Yang S, Wu J, Liu J, Li X, Zhang Q, Pan M, Zhang S (2023b) Exploring sparse visual prompt for cross-domain semantic segmentation. arXiv preprint. [arXiv:2303.09792](https://arxiv.org/abs/2303.09792)
- Yang Y, Jiang P-T, Wang J, Zhang H, Zhao K, Chen J, Li B (2024) Empowering segmentation ability to multi-modal large language models. arXiv preprint. [arXiv:2403.14141](https://arxiv.org/abs/2403.14141)
- Yao H, Hu X, Li X (2022) Enhancing pseudo label quality for semi-supervised domain-generalized medical image segmentation. In: Proceedings of the AAAI conference on artificial intelligence, vol 36, pp 3099–3107
- Yu C, Wang J, Peng C, Gao C, Yu G, Sang N (2018) BiseNet: bilateral segmentation network for real-time semantic segmentation. In: Proceedings of the European conference on computer vision (ECCV), pp 325–341
- Yu F, Chen H, Wang X, Xian W, Chen Y, Liu F, Madhavan V, Darrell T (2020) BDD100K: a diverse driving dataset for heterogeneous multitask learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2636–2645
- Yun S, Han D, Oh SJ, Chun S, Choe J, Yoo Y (2019) Cutmix: regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 6023–6032
- Zhang Y, Yang Q (2021) A survey on multi-task learning. *IEEE Trans Knowl Data Eng* 34(12):5586–5609
- Zhang Q, Zhang J, Liu W, Tao D (2019) Category anchor-guided unsupervised domain adaptation for semantic segmentation. *Adv Neural Inform Proc Syst*. <https://doi.org/10.48550/arXiv.1910.13049>
- Zhang J, Qi L, Shi Y, Gao Y (2020) Generalizable semantic segmentation via model-agnostic learning and target-specific normalization. arXiv preprint. [arXiv:2003.12296](https://arxiv.org/abs/2003.12296)
- Zhang W, Pang J, Chen K, Loy CC (2021a) K-Net: towards unified image segmentation. *Adv Neural Inform Proc Syst* 34:10326–10338
- Zhang J, Zhang Y, Xu X (2021b) ObjectAug: object-level data augmentation for semantic image segmentation. In: 2021 International joint conference on neural networks (IJCNN). IEEE, pp 1–8
- Zhang J, Qi L, Shi Y, Gao Y (2022a) Generalizable model-agnostic semantic segmentation via target-specific normalization. *Pattern Recog* 122:108292. <https://doi.org/10.1016/j.patcog.2021.108292>
- Zhang B, Tian Z, Tang Q, Chu X, Wei X, Shen C (2022b) SegViT: semantic segmentation with plain vision transformers. *Adv Neural Inform Proc Syst* 35:4971–4982
- Zhang Y, Roy S, Lu H, Ricci E, Lathuilière S (2023) Cooperative self-training for multi-target adaptive semantic segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 5604–5613
- Zhang J, Huang J, Jin S, Lu S (2024a) Vision-language models for vision tasks: a survey. *IEEE Trans Pattern Anal Mach Intell*. <https://doi.org/10.48550/arXiv.2304.00685>
- Zhang X, Li J, Chu W, Hai J, Xu R, Yang Y, Guan S, Xu J, Cui P (2024b) On the out-of-distribution generalization of multimodal large language models. arXiv preprint. [arXiv:2402.06599](https://arxiv.org/abs/2402.06599)
- Zhao Y, Zhao N, Lee GH (2022a) Synthetic-to-real domain generalized semantic segmentation for 3D indoor point clouds. arXiv preprint. [arXiv:2212.04668](https://arxiv.org/abs/2212.04668)
- Zhao Y, Zhong Z, Zhao N, Sebe N, Lee GH (2022b) Style-hallucinated dual consistency learning for domain generalized semantic segmentation. In: European conference on computer vision. Springer, Cham, pp 535–552
- Zheng S, Lu J, Zhao H, Zhu X, Luo Z, Wang Y, Fu Y, Feng J, Xiang T, Torr PH (2021) Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 6881–6890
- Zhong Z, Zhao Y, Lee GH, Sebe N (2022) Adversarial style augmentation for domain generalized urban-scene segmentation. *Adv Neural Inform Proc Syst* 35:338–350
- Zhou B, Zhao H, Puig X, Xiao T, Fidler S, Barriuso A, Torralla A (2019) Semantic understanding of scenes through the ade20k dataset. *Int J Comput Vis* 127:302–321
- Zhou Y, Chen H, Li Y, Liu Q, Xu X, Wang S, Yap P-T, Shen D (2021) Multi-task learning for segmentation and classification of tumors in 3d automated breast ultrasound images. *Med Image Anal* 70:101918
- Zhou K, Liu Z, Qiao Y, Xiang T, Loy CC (2022a) Domain generalization: a survey. *IEEE Trans Pattern Anal Mach Intell* 45:4396–4415

- Zhou Z, Qi L, Shi Y (2022b) Generalizable medical image segmentation via random amplitude mixup and domain-specific image restoration. In: European conference on computer vision. Springer, Cham, pp 420–436
- Zhu Y, Kiros R, Zemel R, Salakhutdinov R, Urtasun R, Torralba A, Fidler S (2015) Aligning books and movies: towards story-like visual explanations by watching movies and reading books. In: Proceedings of the IEEE international conference on computer vision, pp 19–27
- Ziegler A, Asano YM (2022) Self-supervised learning of object parts for semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 14502–14511
- Zou Y, Yu Z, Kumar B, Wang J (2018) Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: Proceedings of the european conference on computer vision (ECCV), pp 289–305

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Taki Hasan Rafi¹ · Ratul Mahjabin² · Emon Ghosh³ · Young-Woong Ko¹ · Jeong-Gun Lee¹

✉ Young-Woong Ko
youngwoongKo@hallym.ac.kr

✉ Jeong-Gun Lee
jeonggun.lee@hallym.ac.kr

Taki Hasan Rafi
takihasanrafi@gmail.com

Ratul Mahjabin
ratulmahjabin@iut-dhaka.edu

Emon Ghosh
emonghosh005@gmail.com

¹ Department of Computer Engineering, Hallym University, Chuncheon, South Korea

² Department of Computer Science and Engineering, Islamic University of Technology, Gazipur, Bangladesh

³ Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh