Handling Class-Imbalance for Improved Zero-Shot Domain Generalization

Ahmad Arfeen
ahmadarfeen@iisc.ac.in
Titir Dutta
titird@iisc.ac.in
Soma Biswas
somabiswas@iisc.ac.in

Indian Institute of Science
Bangalore, India

Abstract

Zero-shot domain generalization (ZSDG) simultaneously addresses the challenges of dissimilar distribution and disjoint label-spaces of the training and test data in the context of classification. State-of-the-art ZSDG approaches leverage multiple source domain data and the semantic information of the classes to learn domain-agnostic features for handling both unseen domains and classes. Effective feature learning depends significantly on the training data characteristics, which has been largely overlooked for this task. In this work, we propose to handle one such important challenge, namely class-imbalance for the ZSDG problem. Towards this end, we propose a novel framework, **M**ixing-based **A**daptive **M**argin **C**lassifier **Net**work (MAMC-Net) for handling this real-world challenge. Specifically, it consists of two components, (i) a novel adaptive-margin based semantic classifier for handling the data imbalance in the training data and (ii) a module for determining the mixing ratio when the input domains and classes are mixed, for better domain agnostic class-discrimination. Extensive experiments and analysis performed on multiple large-scale datasets, DomainNet and DomainNet-LS demonstrate the effectiveness of MAMC-Net to address the challenging ZSDG scenario.

1 Introduction

Recent advances in deep learning have brought significant improvements in several computer vision tasks, eg. image classification [14], segmentation [32] etc., but with the underlying assumptions that the test data always belong to the same distribution as the training data, and they share the same label-space. However, in real-life, the test data may belong to any category or domain, and this information is not known a-priori. Research in the directions of domain generalization (DG) [19][28] and zero-shot learning (ZSL) [36][40] address these restrictions individually. DG explores the challenging scenario of classifying data from a completely unknown target domain [22][20] (but identical label-space), by learning a domain-invariant representation, using data collected across multiple source domains. On the other hand, ZSL-algorithms can classify objects from unseen classes (but identical data-distribution), by addressing the knowledge-gap of the seen and unseen categories using their corresponding semantic properties. Motivated by real-world challenges, recently, researchers have started to address zero-shot domain generalization (ZSDG) [27][5], where the test data can belong to an unseen class as well as unseen domain.

ZSDG leverages samples across multiple domains (as in DG) to train the model, and quite often this results in severely imbalanced training data. For example, RGB-images may be easier to collect as compared to drawing *Sketches*. Even within a single domain, it might be much easier to collect data from few classes compared to the others. Thus the training data can have domain as well as class imbalance. The class-imbalance problem has been widely addressed in the context of several applications like image classification [14], object detection [44], etc.

In this work, we address the challenges posed by class-imbalanced training data for the ZSDG task. Towards that goal, we propose a novel framework consisting of a semantic classifier with an adaptive margin to account for the variable number of training data of the different classes. We integrate it with an inter-domain and inter-class mixing network with mixing-ratio prediction ability to ensure that the feature representation is domain-agnostic. We refer to this proposed framework as **M**ixing-based **A**daptive **M**argin **C**lassifier **Net**work (MAMC-Net). Thus the contributions of this work are summarized as follows: (i) We address the real-world problem of class-imbalance for the challenging ZSDG setting, which to the best of our knowledge is the first in literature; (ii) We propose the MAMC-Net framework, where a novel class-specific adaptive margin is learnt to address the imbalance issue; (iii) In addition, we also learn to predict the ratio of a mixed input, which helps in extracting better domain-agnostic features to further boost the performance; (iv) The proposed framework outperforms all the existing approaches and obtains state-of-the-art results on the two large-scale benchmark datasets, DomainNet [31] and DomainNet-LS [27].

2 Related Work

Here, we discuss relevant work in the literature for DG, ZSL, ZSDG and data imbalance.

Domain Generalization (DG): Domain generalization aims to achieve domain invariant feature representations of the data, by training the model on data from multiple domains. In [42], a conditional invariant deep network is trained with adversarial training to obtain the same. The work in [23] extends the adversarial learning process using auto-encoders with Maximum Mean Discrepancy (MMD) loss. A meta-learning based approach is proposed in [18] and a combination of extrinsic and intrinsic supervisions is proposed in [38]. More recently, [8] proposed to learn a domain-representation while adapting to the target domain, and [37] introduces domain-specific batch-normalization to address DG. In [4], a self-supervised approach for learning the input features by solving a jigsaw puzzle created using random patches on the input image is used. [26] proposed to learn the multiple latent-domains in the training data using unsupervised clustering on style-features and cross-entropy based loss on category-features. A modified DG-protocol with single-domain training data using adversarial domain-augmentation is explored in [33].

Zero-Shot Learning (ZSL): ZSL algorithms usually leverage the class attributes or semantic information to bridge the gap between the seen and unseen classes. Such information can be in the form of text [34], attributes [16], knowledge graph (KG) and ontology rules [24] Several ZSL approaches [43][17] learn a mapping function from the image space to the semantic space and then utilize nearest neighbour search to match the query image to one of the class semantics. Recently, generative models have been proposed to generate synthetic samples of the novel classes (not present in the training) using their respective semantic vectors [35][9], which are finally used to train the classification model. Using graph knowledge as the semantic representation is another effective approach for ZSL, as presented in [39].

Zero Shot Domain Generalization (ZSDG): Recently ZSDG has become an active research area, and few algorithms [27][5] have been proposed which report impressive results for this task. In [27], an effective mixing strategy is proposed to generate new synthesized data of unseen categories and unseen domains through inter-class and inter-domain mixing. The state-of-the-art ZSDG approach [5] learns domain independent structure latent embeddings by projecting both the image and the semantic representations to a common latent space. Usually, Word2vec [6]-embeddings have been utilized in these works.

Class Imbalance: This is a well-studied problem in the context of classification [14], object detection [44], etc. One straight-forward approach is to perform over-sampling of minority classes [15][2], but this is usually not very effective [25]. Most of the state-of-the-art approaches address the class imbalance problem by modifying the classifier. Towards this end, [12] learns the classifier to maximally spread out in the embedding space. In [3], the classifiers are learnt to obtain broader margin for minority classes compared to the others.

3 Problem Definition and Motivation

First, we introduce the different notations used and also discuss the motivation for this work. We denote the training data as $\mathcal{D}_{tr} = \{(x_i, y_i, d_i)\}_{i=1}^{n}$, containing n-number of samples. Here, x_i represents the i^{th} sample, from class y_i and domain d_i. The corresponding seen-label space and the set of training-domains are represented as $\mathcal{Y}_{tr} = \{y_1, \ldots, y_i, \ldots\}$, and $\mathcal{S}_{tr} = \{d_1, \ldots, d_i, \ldots\}$, respectively. The goal of ZSDG [27][5] is to classify the test samples from $\mathcal{D}_{te} = \{x_{te}\}$, where the test label-space $\mathcal{Y}_{te}$ and domain-set $\mathcal{S}_{te}$ corresponding to $\mathcal{D}_{te}$ are strictly non-overlapping with the training sets, ie.: $\mathcal{Y}_{tr} \cap \mathcal{Y}_{te} = \phi$ and $\mathcal{S}_{tr} \cap \mathcal{S}_{te} = \phi$.
Motivation: It is well known that training data characteristics like data imbalance adversely affects the final performance for several tasks like image classification, object detection, etc. and addressing them can significantly boost the results. But these factors have been largely overlooked in tasks like ZSDG. Analyzing the large-scale benchmark dataset for DG and ZSDG, namely DomainNet [31], we observe from Figure 1 (a,b), that significant class imbalance exists across multiple domains, which justifies the effort to address the class-imbalance problem for the ZSDG task. In Figure 1 (c), we observe from the t-SNE plot that the samples from the 5-majority (spreadsheet, table, tree, whale and bird) and 5-minority (dresser, calender, ceiling-fan, saw and line) classes cluster in the feature-space, justifying the effectiveness of the proposed MAMC-Net. Next, we discuss the proposed framework.

4 Mixing-based Adaptive Margin Classifier Network

The proposed MAMC-Net constitutes two main modules, namely (i) Domain-Agnostic Representation Module and (ii) Class-Imbalance Handling Module, on top of a base model. The base model can potentially be any semantic-classifier based existing ZSDG network. We demonstrate that the integration of the two proposed modules with one such base model enhances its performance, by addressing the class-imbalance in the training set. Here, we

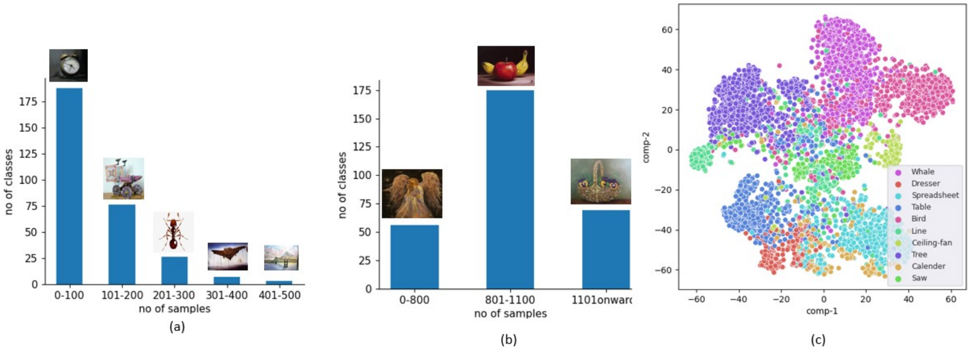

Figure 1: Severe class imbalance is observed from the data distribution of (a) Infograph domain; (b) all the source domains combined for DomainNet dataset. (c) t-SNE plot of the MAMC-Net feature space for 5 majority and 5 minority training classes.

consider CuMix [27] to be the base model, which we briefly describe below.

Base Model: Our base network, CuMix [27], is an inter-domain and intra-domain mixing based network, and the first to address ZSDG. It consists of a mixing module $\mathcal{M}$, a feature-extractor $\mathcal{F}$ and a classifier $\mathcal{C}$. The mixing in this network is performed on a triplet set $\mathcal{T} = \{x_i, x_j, x_k\}$, such that the sample-pair $\{x_i, x_j\}$ are from different domains (inter-domain samples), and $\{x_i, x_k\}$ are from same domain (intra-domain samples), but belonging to different categories (inter-class samples). Thus, the mixed sample is

$$x_{mixed} = \mathcal{M}(x_i, x_j, x_k, \delta) = \delta x_i + (1-\delta)[\xi x_j + (1-\xi)x_k] \quad (1)$$

where, δ is the mixing co-efficient and ξ is sampled from Binomial distribution, generating x_{mixed} as either a cross-domain sample (for $\xi = 1$), or a cross-category sample ($\xi = 0$). The rest of the components of the network $\mathcal{F}$ and $\mathcal{C}$ are learned using x_{mixed}.

The classifier module $\mathcal{C}$ in CuMix consists of (1) a semantic projector $\mathcal{P}_{sem}$ to transform the feature-representation $\mathcal{F}(x_{mixed})$ to the semantic space, and (2) a set of fixed semantic weights, $\mathcal{W}_{sem} = [w_1, \ldots, w_c, \ldots, w_{|\mathcal{Y}_{tr}|}]$, which is the collection of semantic information (w_c) in the form of word-embedding of the category-names for each class $y_c \in \mathcal{Y}_{tr}$. These weights are utilized to compute the probability that x_{mixed} belongs to class with label y_c as, $prob(x_{mixed} \in y_c) = \frac{exp(w_c * \mathcal{F}(x_{mixed}))}{\sum_{l \in \mathcal{Y}_{tr}} exp(w_l * \mathcal{F}(x_{mixed}))}$. Here we overload the notation y_c to also denote the class with that label. The network is trained in an end-to-end manner through minimizing a mix-up based cross-entropy loss based on these probabilities as,

$$\begin{aligned}\mathcal{L}_{mixed-CE} = -\frac{1}{|\mathcal{T}|}[\delta \log prob(x_{mixed} \in y_i) \\ + (1-\delta)[\xi \log prob(x_{mixed} \in y_j) \\ + (1-\xi) \log prob(x_{mixed} \in y_k)]] \quad (2)\end{aligned}$$

In addition, the model also minimizes a semantic loss function $\mathcal{L}_{sem}$, which is the cross-

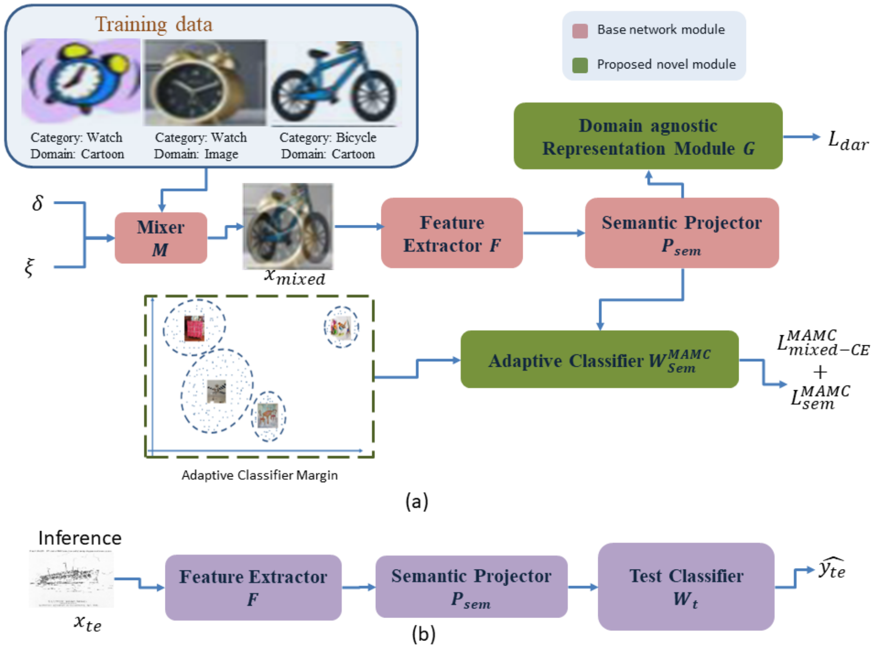

Figure 2: The proposed MAMC-Net architecture for (a) training, (b) inference.

entropy loss for the raw-samples $x_i \in \mathcal{D}_{tr}$ as

$$\mathcal{L}_{sem} = -\frac{1}{n} \sum_{y_c \in \mathcal{Y}_{tr}} \log prob(x_i \in y_c) \tag{3}$$

Now, we describe the proposed novel modules in MAMC-Net in details. A block diagram of the proposed framework is shown in Figure 2.

1) Domain Agnostic Representation Module: Inspired by [30], we introduce a module $\mathcal{G}$ to learn the mixing proportion of the different classes for each x_{mixed}. This will enable the model to forget the domain specific characteristics present in x_{mixed}, and to learn only the features specific to the category. This serves the domain-agnostic representation learning goal, as well as enhances the class-discriminability of the model. The corresponding loss is:

$$\mathcal{L}_{dar} = -\frac{1}{|\mathcal{T}|} \sum_{x_{mixed} \in \mathcal{T}} \eta_{mixed} * \log(v_{mixed}); \qquad v_{mixed} = \text{soft-max}(\mathcal{G}(x_{mixed})) \tag{4}$$

where, η_{mixed} is the mixed-label information of x_{mixed}, computed on the basis of the one-hot representation (η_i, η_j and η_k) of component classes as $\eta_{mixed} = \delta\eta_i + (1-\delta)[\xi\eta_j + (1-\xi)\eta_k]$. This loss improves the domain-invariant representation of the feature-extractor $\mathcal{F}$.

2) Adaptive Margin Classifier Module: Here, we explain the novel adaptive margin classifier, which can account for the class imbalance in the training data, while computing discriminating class-boundaries for the training classes. The confusion for a class depends on its variance, and during training, the classifier learns this spread or boundary about the class from the training data. When there is class imbalance, for minority classes, confusion arises since adequate data for this learning is not available, and thus a larger classifier margin tries to compensate for it. In ZSDG, though the test classes are different from the training ones, they are classified using their relatedness to the seen ones. Thus, if the training (seen) classes are improperly classified due to inadequate training examples in few classes or other reasons, it will in turn adversely affect the test classification.

Several state-of-the-art class-imbalance handling techniques for applications like image classification aim to modify the classifier [3][12]. In contrast, the classifiers ($\mathcal{W}_{sem}$) in ZSDG are usually fixed to the semantic representations (word-embeddings of class-names) to account for the unseen classes that will be encountered during testing. As explained earlier, for the minority classes, lesser number of training samples may result in increased confusion with the neighbouring classes. To account for this, we propose to incorporate an adaptive margin to each of the fixed semantic classifiers, with the objective that the classifiers sampled from this margin should be able to classify samples from the corresponding class correctly. Since the class probabilities are obtained by the cosine similarity between the feature and the perturbed classifier, this mitigates the adverse affect of lesser number of samples. Another advantage is that this class-specific margins can be learnt in an end-to-end-manner. Specifically, this margin is computed as a scaled Gaussian noise, where the scaling factor is learned during training. Specifically, the margin-factor for class $y_c \in \mathcal{Y}_{tr}$ is defined as,

$$\Delta_c = \lambda_c \mathcal{N}(0,1) \tag{5}$$

where $\mathcal{N}(0,1)$ is a zero-mean unit-variance Gaussian noise and λ_c is the learnable scaling factor. We initialize λ_c as per the number of samples n_c of class y_c in the training set as, $\lambda_c = (1 - \frac{n_c}{n})$. For minority classes, $n_c << n$, and thus λ_c has a higher value, resulting in a wider Δ_c around the semantic embedding of the category name of y_c. For majority classes, this margin is less. Throughout the training, the weights of the proposed adaptive classifier w_c^{MAMC} for class c is randomly sampled following $w_c^{MAMC} \sim (w_c + \Delta_c)$. Using this modification, we now compute the probability of x_{mixed} to belong to class c as, $\mathcal{P}(x_{mixed} \in y_c, \lambda_c) = \frac{exp(w_c^{MAMC} * \mathcal{F}(\mathbf{x}_{mixed}))}{\sum_{l \in \mathcal{Y}_{tr}} exp(w_l^{MAMC} * \mathcal{F}(\mathbf{x}_{mixed}))}$, and the corresponding cross-entropy loss as,

$$\begin{aligned} \mathcal{L}_{mixed-CE}^{MAMC} = -\frac{1}{|\mathcal{T}|}[\delta \log \mathcal{P}(x_{mixed} \in y_i, \lambda_i) \\ + (1-\delta)[\xi \log \mathcal{P}(x_{mixed} \in y_j, \lambda_j) \\ + (1-\xi) \log \mathcal{P}(x_{mixed} \in y_k, \lambda_k)]] \end{aligned} \tag{6}$$

Now, we discuss the complete training process followed in this work.

Complete Training and Inference Methodology: For effectively training MAMC-Net, we minimize a combination of three loss-components: (1) Mix-up based cross-entropy loss $\mathcal{L}_{mixed-CE}^{MAMC}$ with the adaptive classifier; (2) Domain agnostic representation loss $\mathcal{L}_{dar}$; (3) Modified semantic loss (eq. (3)) given by: $\mathcal{L}_{sem}^{MAMC} = -\frac{1}{n} \sum_{y_c \in \mathcal{Y}_{tr}} \log \mathcal{P}(x_{mixed} \in y_c, \lambda_c)$. Thus, the combined loss function is given by

$$\mathcal{L} = \kappa_1 \mathcal{L}_{sem}^{MAMC} + \kappa_2 \mathcal{L}_{dar} + \kappa_3 \mathcal{L}_{mixed-CE}^{MAMC} \tag{7}$$

where, κ_1, κ_2 and κ_3 are the hyper-parameters. The model is trained end-to-end to optimize the model components, $\mathcal{F}$, $\mathcal{P}_{sem}$, $\mathcal{G}$ and the margins Δ_c in the adaptive classifier $\mathcal{W}_{sem}^{MAMC}$ for handling class-imbalanced ZSDG task.

After training, the obtained feature extractor $\mathcal{F}$ can robustly classify samples from unseen domain(s), as well as from unseen class(es). During inference, we use the semantic information from these unseen classes in $\mathcal{Y}_{te}$ as the classifiers, i.e. the classifier for t^{th} unknown class is w_t, contains the semantic information in the form of the word-embeddings of t^{th}-class's name. The test sample $x_{te} \in \mathcal{D}_{te}$ is passed through the feature extractor $\mathcal{F}$, and the corresponding class is predicted as

$$\hat{y}_{te} = \underset{t \in \mathcal{Y}_{te}}{\arg\max} \, [w_t * \mathcal{F}(x_{te})] \tag{8}$$

Now, we will discuss the experiments performed to evaluate the proposed MAMC-Net.

5 Experiments and Analysis

We experimented on two large-scale datasets, namely DomainNet [31] and DomainNet-LS [27]. **DomainNet [31]** is the benchmark dataset for ZSDG [27][6] and comprises of 6-domains consisting of 345-categories in each domain. The domains are real, painting, sketch, quickdraw, info-graph and clip-art. Following the standard ZSDG-protocol [27], we use 300-categories for training, and the remaining for evaluation. 5-domains are used during training, and the remaining domain is used as the unseen domain for testing. However, image-domain is never used as an unseen target domain, as the model is pre-trained on the ImageNet [7] data. **DomainNet-LS** is a modified DomainNet dataset [31], where the training and test domains are pre-defined. Here, the class-wise seen / unseen data split remains unchanged, but only the two domains, real and painting are used for training. The remaining 4-domains are used for evaluation. This domain-split becomes more challenging due to the presence of unseen domains, like sketch or quickdraw, for which no related training domains are present. This dataset was first proposed in [27], and later used in [6] for ZSDG.

Baselines: Following [27], we compare MAMC-Net with several existing ZSDG methods in literature, such as: (1) **ZSDG:** We provide comparison with SOTA methods, namely CuMix [27] and SLE-Net [6]. We utilized CuMix [27] as the base network and explained it in Section 4. SLE-Net [6] is the more recent approach for ZSDG. (2) **ZSL:** We report the performance of few ZSL methods, such as DEVISE [10], ALE [1] and SPNET [41] under the ZSDG scenario. We train these models on the training data from multiple domains, and perform the inference on data from an unknown domain. (3) **Extending DG for ZSL:** We also experiment with standard DG methods, such as DANN [11] and EpiFCR [21], and use them in addition with the ZSL-methods to perform classification under ZSDG protocol. Thus, the baselines selected to benchmark the performance of MAMC-Net is quite exhaustive.

Implementation details : MAMC-Net is implemented in Pytorch [29], using a single Nvidia RTX A5000 GPU. We use Resnet-50 [13], pre-trained on ImageNet as the backbone (same as CuMix [27]). We report the results using standard top-1 average per-class accuracy. Following [27], the semantic representation of the classes are the L2-normalized word-embeddings from Google news corpus word2vec [6] model. SGD optimizer with momentum = 0.9, an initial learning rate of 10^{-3}, with a multi-step scheduler with scale factor of 0.1 at each

Table 1: ZSDG results on DomainNet dataset - per class accuracy (%).

Method		Target Domain					Avg
DG	ZSL	painting	inforgraph	quickdraw	sketch	clipart	
-	DEVISE [10]	17.6	11.7	6.1	16.7	20.1	14.4
	ALE [1]	20.2	12.7	6.8	18.5	22.7	16.2
	SPNet [41]	23.8	16.9	8.2	21.8	26.0	19.4
DANN [11]	DEVISE [10]	16.4	10.4	7.1	15.1	20.5	13.9
	ALE [1]	19.7	12.5	7.4	17.9	21.2	15.7
	SPNet [41]	24.1	15.8	8.4	21.3	25.9	19.1
EpiFCR [21]	DEVISE [10]	19.3	13.9	7.3	17.2	21.6	15.9
	ALE [1]	21.4	14.1	7.8	20.9	23.2	17.5
	SPNet [41]	24.6	16.7	9.2	23.2	26.4	20.0
CuMix [27]		25.5	17.8	9.9	22.6	27.6	20.7
CuMix (our implementation)		25.2	17.1	9.3	22.1	26.5	20.0
SLE-Net [5] (SOTA)		26.6	18.4	11.5	25.2	27.8	21.9
MAMC-Net (Ours)		**27.3**	**19.5**	**12.1**	**26.0**	**28.8**	**22.7**

Table 2: ZSDG results on DomainNet dataset - top-1 standard accuracy (%).

Method	Target Domain					Avg
	painting	inforgraph	quickdraw	sketch	clipart	
CuMix [27]	27.6	16.3	9.7	25.9	27.8	21.5
SLE-Net [5] (SOTA)	28.8	17.6	11.5	26.3	29.1	22.7
MAMC-Net (Ours)	**29.2**	**18.8**	**12.2**	**27.4**	**30.0**	**23.5**

Table 3: ZSDG results on DomainNet-LS dataset.

Method	Target Domain				Avg.
	quickdraw	sketch	inforgraph	clipart	
SPNet [41]	4.8	17.3	14.1	21.5	14.4
Epi-FCR [21] + SPNet [41]	5.6	18.7	14.9	22.5	15.4
CuMix [27]	5.5	19.7	17.1	23.7	16.5
SLE-Net [5] (SOTA)	7.2	20.5	16	**24**	16.9
MAMC-Net (Ours)	**8.2**	**21.2**	**17.6**	23.6	**17.7**

scheduling step is used for learning, on a batch-size of 100. The mixing ratio δ in our experiments are sampled from a beta-distribution $\beta(m,n)$, where $m = n$, and m is changed at regular intervals to ensure fair contribution for all component samples. All the hyper-parameters $k_1 = k_2 = k_3 = 1$, since the starting values for all three loss components in eqn. 7 are in comparable range. The average accuracy results over 10 trials for each target domain are reported in this work.

Results on DomainNet dataset: We summarize the ZSDG results on DomainNet in Table 1 in terms of per-class accuracy. The results for other algorithms are directly taken from [5]. We also include results of CuMix using our implementation (since this is our base network), which is similar to the reported ones [27], except for clipart. We observe that the performance of the ZSL-methods decrease when fused with DG method DANN [11] for almost all target domains, while EpiFCR [21] improves the ZSL methods marginally. These results indicate the need for developing dedicated methods to generalize across classes and domains. This is evident from the performance of CuMix and SLE-Net, which perform much better compared to the previous ones. However, MAMC-Net outperforms all these methods by a significant margin. Even in terms of top-1 standard accuracy, MAMC-Net outperforms both CuMix and SLE-Net as observed in Table 2. This clearly justifies the effectiveness of the

Table 4: Ablation study of proposed MAMC-Net on DomainNet.

Model Variants	quickdraw	painting
Base Model	9.2	25.2
Base model + $\mathcal{L}_{dar}$	10.5	26.4
Base model + $\mathcal{L}_{dar}$ + $\mathcal{L}_{fix-margin}$	11.1	26.3
MAMC-Net	12.1	27.3

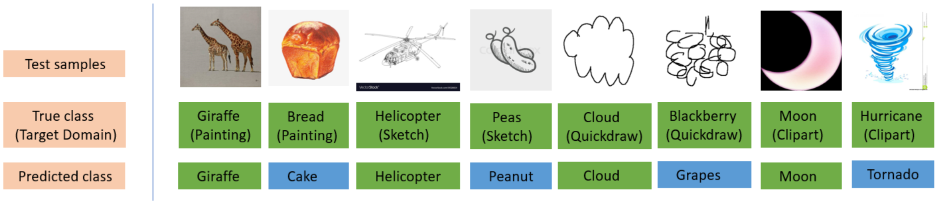

Figure 3: MAMC-Net predictions for few test samples from DomainNet dataset.

proposed method for ZSDG task. We also present sample prediction results on this dataset using MAMC-Net in Figure 3. The correct and incorrect predictions are highlighted with *green* and *blue*, respectively. We observe that most of the incorrect predictions are semantically related to the query images (eg. *bread* is classified as *cake*, *blackberry* is classified as *grapes*). This indicates MAMC-Net's effectiveness to fuse the semantic relevance between categories while learning the feature-space.

Results on DomainNet-LS dataset: The results for DomainNet-LS dataset are reported in Table 3. The performance of the ZSL method, SPNet on the four targets, and its performance when combined with the DG-method Epi-FCR shows that this fusion provides marginal improvement for ZSDG task. In comparison, the ZSDG methods perform much better. However, the proposed MAMC-Net outperforms all the other ZSDG methods, CuMix and SLE-Net by a significant margin.

Ablation Study: We perform an ablation study for MAMC-Net (Table 4) to understand the effectiveness of each component. We perform this study on DomainNet, with Quickdraw (ambigous and roughly drawn sketches, thus difficult to predict) and Painting (visually similar to RGB, thus easily predictable domain) as the target domains, which create robust test-conditions for the model. We perform experiments with different variations of the model, such as: (1) *Base Model:* gives the results of our base network CuMix [21]; (2) *Base model* + $\mathcal{L}_{dar}$: Here we use only the domain-agnostic loss with the fixed semantic classifier $\mathcal{W}_{sem}$; (3) *Base model* + $\mathcal{L}_{dar}$+ $\mathcal{L}_{fix-margin}$: Here, we use a fixed margin ($\Delta_c = (1 - \frac{n_c}{n})\mathcal{N}(0,1)$) based classifier, instead of a learnable one; (4) *MAMC-Net:* This includes all the proposed modules. We observe that $\mathcal{L}_{dar}$ provides a significant boost to the base network, and the fix-margin classifier further improves the results (specially for quickdraw), justifying such margin-based classifier design. MAMC-Net outperforms all the other model-variants. Thus each of the proposed modules contributes to the good performance of the proposed framework.

6 Conclusion

In this work, we proposed a novel framework, MAMC-Net for handling the class imbalance problem in the context of zero-shot domain generalization. In addition to a domain-agnostic representation learning loss, we also introduced a novel learnable margin to the existing semantic classifier of a base ZSDG-network. We also presented extensive experiments and analysis across two large-scale datasets to demonstrate the effectiveness of this method. Here, we have addressed the class-imbalance in the training data, which can be extended for handling domain imbalance, resulting in further improvement in the ZSDG performance.

7 Acknowledgement

This work is partly supported through a research grant from SERB, Department of Science and Technology, Govt. of India.

References

[1] Zeynep Akata, Florent Perronnin, Zaid Harchaoui, and Cordelia Schmid. Label-embedding for attribute-based classification. In *CVPR*, 2013.

[2] Adnan Amin, Sajid Anwar, Awais Adnan, Muhammad Nawaz, Newton Howard, Junaid Qadir, Ahmad Hawalah, and Amir Hussain. Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study. *IEEE Access*, 4, 2016.

[3] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. *NeurIPS*, 2019.

[4] F. M. Carlucci, A. D'Innocente, S. Bucci, B. Caputo, and T. Tommasi. Domain generalization by solving jigsaw puzzles. In *CVPR*, 2019.

[5] Shivam Chandhok, Sanath Narayan, Hisham Cholakkal, Rao Muhammad Anwer, Vineeth N Balasubramanian, Fahad Shahbaz Khan, and Ling Shao. Structured latent embeddings for recognizing unseen classes in unseen domains. *BMVC*, 2021.

[6] Kenneth Ward Church. Word2vec. *Natural Language Engineering*, 23, 2017.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009.

[8] Abhimanyu Dubey, Vignesh Ramanathan, Alex Pentland, and Dhruv Mahajan. Adaptive methods for real-world domain generalization. In *CVPR*, 2021.

[9] Rafael Felix, Ian Reid, Gustavo Carneiro, et al. Multi-modal cycle-consistent generalized zero-shot learning. In *ECCV*, 2018.

[10] Andrea Frome, Greg S Corrado, Jon Shlens, Samy Bengio, Jeff Dean, Marc'Aurelio Ranzato, and Tomas Mikolov. Devise: A deep visual-semantic embedding model. *NeurIPS*, 2013.

[11] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *machine learning research*, 17(1), 2016.

[12] Munawar Hayat, Salman Khan, Syed Waqas Zamir, Jianbing Shen, and Ling Shao. Gaussian affinity for max-margin class imbalanced learning. In *ICCV*, 2019.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[14] Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang. Learning deep representation for imbalanced classification. In *CVPR*, 2016.

[15] Bartosz Krawczyk. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4), 2016.

[16] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Attribute-based classification for zero-shot visual object categorization. *PAMI*, 36, 2013.

[17] Jimmy Lei Ba, Kevin Swersky, Sanja Fidler, et al. Predicting deep zero-shot convolutional neural networks using textual descriptions. In *ICCV*, 2015.

[18] D. Li, Y. Yang, Y. Z. Song, and T. M. Hospedales. Learning to generalize: meta-learning for domain generalization. In *AAAI*, 2018.

[19] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In *ICCV*, 2017.

[20] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Learning to generalize: Meta-learning for domain generalization. In *AAAI*, 2018.

[21] Da Li, Jianshu Zhang, Yongxin Yang, Cong Liu, Yi-Zhe Song, and Timothy M Hospedales. Episodic training for domain generalization. In *ICCV*, 2019.

[22] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In *CVPR*, 2018.

[23] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In *CVPR*, 2018.

[24] Juan Li, Ruoxu Wang, Ningyu Zhang, Wen Zhang, Fan Yang, and Huajun Chen. Logic-guided semantic representation learning for zero-shot relation classification. *arXiv preprint arXiv:2010.16068*, 2020.

[25] Li Li, Kaiyi Zhao, Ruizhi Sun, Jiangzhang Gan, Gang Yuan, and Tong Liu. Parameter-free extreme learning machine for imbalanced classification. *NPL*, 52(3), 2020.

[26] T. Maatsura and T. Harada. Domain generalization using a mixture of multiple latent domains. In *AAAI*, 2020.

[27] Massimiliano Mancini, Zeynep Akata, Elisa Ricci, and Barbara Caputo. Towards recognizing unseen categories in unseen domains. In *ECCV*, 2020.

[28] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *ICML*, 2013.

[29] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *NeurIPS*, 2019.

[30] Soumava Paul, Titir Dutta, and Soma Biswas. Universal cross-domain retrieval: Generalizing across classes and domains. In *ICCV*, 2021.

[31] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *ICCV*, 2019.

[32] Dzung L Pham, Chenyang Xu, and Jerry L Prince. Current methods in medical image segmentation. *Annual review of biomedical engineering*, 2, 2000.

[33] Fengchun Qiao, Long Zhao, and Xi Peng. Learning to learn single domain generalization. In *CVPR*, 2020.

[34] Ruizhi Qiao, Lingqiao Liu, Chunhua Shen, and Anton Van Den Hengel. Less is more: zero-shot learning from online textual documents with noise suppression. In *CVPR*, 2016.

[35] Pengda Qin, Xin Wang, Wenhu Chen, Chunyun Zhang, Weiran Xu, and William Yang Wang. Generative adversarial zero-shot relational learning for knowledge graphs. In *AAAI*, 2020.

[36] Bernardino Romera-Paredes and Philip Torr. An embarrassingly simple approach to zero-shot learning. In *ICML*, 2015.

[37] Seonguk Seo, Yumin Suh, Dongwan Kim, Geeho Kim, Jongwoo Han, and Bohyung Han. Learning to optimize domain specific normalization for domain generalization. In *ECCV*, 2020.

[38] Shujun Wang, Lequan Yu, Caizi Li, Chi-Wing Fu, and Pheng-Ann Heng. Learning from extrinsic and intrinsic supervisions for domain generalization. In *ECCV*, 2020.

[39] Xiaolong Wang, Yufei Ye, and Abhinav Gupta. Zero-shot recognition via semantic embeddings and knowledge graphs. In *CVPR*, 2018.

[40] Yongqin Xian, Bernt Schiele, and Zeynep Akata. Zero-shot learning-the good, the bad and the ugly. In *CVPR*, 2017.

[41] Yongqin Xian, Subhabrata Choudhury, Yang He, Bernt Schiele, and Zeynep Akata. Semantic projection network for zero-and few-label semantic segmentation. In *CVPR*, 2019.

[42] L. Ya, X. Tian, M. Gong, Y. Liu, T. Liu, K. Zhang, and D. Tao. Deep domain generalization via conditional invariant adversarial networks. In *ECCV*, 2018.

[43] Yongxin Yang and Timothy M Hospedales. A unified perspective on multi-domain and multi-task learning. *arXiv preprint arXiv:1412.7489*, 2014.

[44] Peng Zhou, Bingbing Ni, Cong Geng, Jianguo Hu, and Yi Xu. Scale-transferrable object detection. In *CVPR*, 2018.