Adversarial Invariant Feature Learning with Accuracy Constraint for Domain Generalization

带有准确性约束的对抗不变特征学习用于领域泛化

Kei Akuzawa [1]☒ , Yusuke Iwasawa [1] , and Yutaka Matsuo [1]

秋泽圭 [1]☒ ，岩泽佑介 [1] ，松尾丰 [1]

School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo

东京大学工程学院，东京都文京区本乡 7-3-1

113-8656, Japan. {akuzawa-kei, iwasawa, matsuo}@weblab.t.u-tokyo.ac.jp

113-8656，日本。{akuzawa-kei, iwasawa, matsuo}@weblab.t.u-tokyo.ac.jp

Abstract. Learning domain-invariant representation is a dominant approach for domain generalization (DG), where we need to build a classifier that is robust toward domain shifts. However, previous domain-invariance-based methods overlooked the underlying dependency of classes on domains, which is responsible for the trade-off between classification accuracy and domain invariance. Because the primary purpose of DG is to classify unseen domains rather than the invariance itself, the improvement of the invariance can negatively affect DG performance under this trade-off. To overcome the problem, this study first expands the analysis of the tradeoff by Xie et. al. [33], and provides the notion of accuracy-constrained domain invariance, which means the maximum domain invariance within a range that does not interfere with accuracy. We then propose a novel method adversarial feature learning with accuracy constraint (AFLAC), which explicitly leads to that invariance on adversarial training. Empirical validations show that the performance of AFLAC is superior to that of domain-invariance-based methods on both synthetic and three real-world datasets, supporting the importance of considering the dependency and the efficacy of the proposed method.

摘要: 学习领域不变表示是领域泛化 (DG) 的主流方法，在这种方法中，我们需要构建一个对领域变化具有鲁棒性的分类器。然而，以前的基于领域不变性的研究忽视了类别与领域之间的潜在依赖关系，这种依赖关系是分类准确性与领域不变性之间权衡的根源。由于 DG 的主要目的是对未见领域进行分类，而不是不变性本身，因此在这一权衡下，不变性的提升可能会对 DG 性能产生负面影响。为了解决这个问题，本研究首先扩展了 Xie 等人 [33] 对权衡的分析，并提供了准确性约束的领域不变性概念，意味着在不干扰准确性的范围内的最大领域不变性。然后，我们提出了一种新方法——带有准确性约束的对抗特征学习 (AFLAC)，该方法明确地在对抗训练中实现了这种不变性。实证验证表明，AFLAC 在合成数据集和三个真实世界数据集上的表现优于基于领域不变性的方法，支持了考虑依赖关系和所提方法有效性的重要性。

Keywords: Invariant Feature Learning - Adversarial Training - Domain Generalization - Transfer Learning

关键词: 不变特征学习 - 对抗训练 - 领域泛化 - 迁移学习

1 Introduction

1 引言

In supervised learning we typically assume that samples are obtained from the same distribution in training and testing; however, because this assumption does not hold in many practical situations it reduces the classification accuracy for the test data [30]. This motivates research into domain adaptation (DA) [9] and domain generalization (DG) [3]. DA methods operate in the setting where we have access to source and (either labeled or unlabeled) target domain data during training, and run some adaptation step to compensate for the domain shift. DG addresses the harder setting, where we have labeled data from several source domains and collectively exploit them such that the trained system generalizes to target domain data without requiring any access to them. Such challenges arise in many applications, e.g., hand-writing recognition (where domain shifts are induced by users, [28]), robust speech recognition (by acoustic conditions, [29]), and wearable sensor data interpretation (by users, [7]).

在监督学习中，我们通常假设训练和测试样本来自相同的分布；然而，由于这一假设在许多实际情况下并不成立，因此降低了测试数据的分类准确性 [30]。这激励了对领域适应 (DA) [9] 和领域泛化 (DG) [3] 的研究。DA 方法在训练期间可以访问源域和 (标记或未标记的) 目标域数据的情况下运行，并进行一些适应步骤以补偿领域转移。DG 处理更困难的情况，在这种情况下，我们拥有来自多个源域的标记数据，并共同利用它们，以便训练的系统能够泛化到目标域数据，而无需访问这些数据。这种挑战出现在

许多应用中，例如手写识别 (用户引起的领域转移，[28])、鲁棒语音识别 (由声学条件引起，[29]) 和可穿戴传感器数据解释 (由用户引起，[7])。

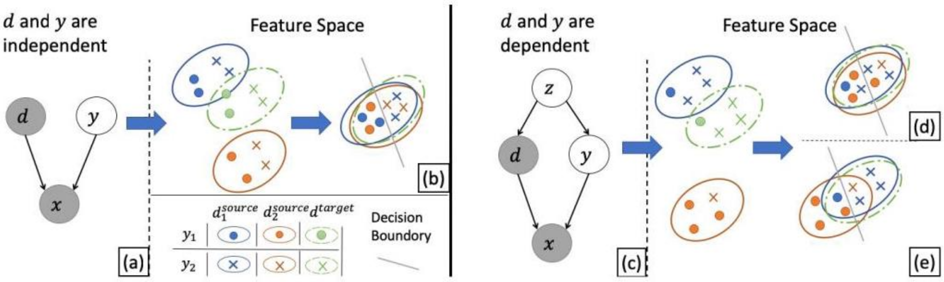


Fig. 1. Explanation of domain-class dependency and the induced trade-off. (a) When the domain and the class are independent, (b) domain invariance and classification accuracy can be optimized at the same time, and the invariance prevents the classifier from overfitting to source domains. (c) When they are dependent, a trade-off exists between these two: (d) optimal classification accuracy cannot be achieved when perfect invariance is achieved, and (e) vice versa. We propose a method to lead explicitly to (e) rather than (d), because the primary purpose for domain generalization is classification, not domain-invariance itself.

图 1. 领域-类别依赖关系及其引发的权衡解释。(a) 当领域和类别独立时，(b) 领域不变性和分类准确性可以同时优化，并且不变性防止分类器对源域过拟合。(c) 当它们相互依赖时，这两者之间存在权衡:(d) 当实现完美不变性时，无法达到最佳分类准确性，而 (e) 则相反。我们提出了一种方法，明确导向 (e) 而不是 (d)，因为领域泛化的主要目的是分类，而不是领域不变性本身。

This paper considers DG under the situation where domain d and class labels y are statistically dependent owing to some common latent factor z (Figure 1-(c)), which we referred to as domain-class dependency. For example, the WISDM Activity Prediction dataset [16], where classes and domains correspond to activities and wearable device users, exhibits this dependency because of the (1) data characteristics: some activities (jogging and climbing stairs) are strenuous to the extent that some unathletic subjects avoided them, and (2) data-collection errors: other activities were added only after the study began and the initial subjects could not perform them. Note that the dependency is common in real-world datasets and a similar setting has been investigated in DA studies [36 12], but most prior DG studies overlooked the dependency; moreover, we need to follow a approach separate from DA because DG methods cannot require any access to target data, as we discuss further in Sec. 2.2

本文考虑了在领域 d 和类别标签 y 由于某些共同潜在因素 z 而统计依赖的情况下的领域泛化 (DG)(图 1-(c))，我们称之为领域-类别依赖。例如，WISDM 活动预测数据集 [16]，其中类别和领域分别对应于活动和可穿戴设备用户，表现出这种依赖性，原因在于 (1) 数据特征: 某些活动 (慢跑和爬楼梯) 是如此剧烈，以至于一些不擅长运动的受试者避免了这些活动，以及 (2) 数据收集错误: 其他活动仅在研究开始后才被添加，而最初的受试者无法执行这些活动。请注意，这种依赖性在现实世界的数据集中是普遍存在的，类似的设置在领域适应 (DA) 研究中也有探讨 [36 12]，但大多数先前的 DG 研究忽视了这种依赖性；此外，我们需要遵循一种与 DA 分开的方法，因为 DG 方法不可以要求访问目标数据，正如我们在第 2.2 节中进一步讨论的那样。

Most prior DG methods utilize invariant feature learning (IFL) [27,7,10,33], which can be negatively affected by the dependency. IFL attempts to learn latent representation h from input data x which is invariant to domains d , or match multiple source domain distributions in feature space. When source and target domains have some common structure (see, 27]), matching multiple source domains leads to match source and target ones and thereby prevent the classifier from overfitting to source domains (Figure 1-(b)). However, under the dependency, merely imposing the perfect domain invariance (which means h and d are independent) adversely affects the classification accuracy as pointed out by Xie et al. [33] and illustrated in Figure 1 Intuitively speaking, since y contains information about d under the dependency, encoding information about d into h helps to predict y ; however, IFL attempts to remove all domain information from h , which causes the trade-off. Although that trade-off occurs in source domains (because we use only source data during optimization), it can also negatively affect the classification performance for target domains. For example, if the target domain has characteristics similar (or same as an extreme case) to those of a certain source domain, giving priority to domain invariance obviously interferes with the DG performance (Figure 1-(d)).

大多数先前的领域泛化 (DG) 方法利用不变特征学习 (IFL)[27,7,10,33]，而这种方法可能会受到依赖性的负面影响。IFL 试图从输入数据 x 中学习潜在表示 h ，该表示对领域 d 是不变的，或者在特征空间

中匹配多个源领域分布。当源领域和目标领域具有某种共同结构时 (见 27)，匹配多个源领域会导致匹配源领域和目标领域，从而防止分类器对源领域的过拟合 (图 1-(b))。然而，在依赖性下，仅仅施加完美的领域不变性 (这意味着 h 和 d 是独立的) 会对分类准确性产生不利影响，正如 Xie 等人所指出的 [33] 并在图 1 中说明的。直观地说，由于 y 在依赖性下包含关于 d 的信息，将关于 d 的信息编码到 h 中有助于预测 y；然而，IFL 试图从 h 中去除所有领域信息，这导致了权衡。尽管这种权衡发生在源领域 (因为我们在优化过程中仅使用源数据)，但它也可能对目标领域的分类性能产生负面影响。例如，如果目标领域的特征与某个源领域的特征相似 (或在极端情况下相同)，优先考虑领域不变性显然会干扰 DG 性能 (图 1-(d))。

In this paper, considering that prioritizing domain invariance under the tradeoff can negatively affect the DG performance, we propose to maximize domain invariance within a range that does not interfere with the classification accuracy (Figure 1-(e)). We first expand the analysis by [33] about domain adversarial nets (DAN), a well-used IFL method, and derive Theorem 1 and 2 which show the conditions under which domain invariance harms the classification accuracy. In Theorem 3 we show that accuracy-constrained domain invariance, which we define as the maximum $H(d \mid h)$ (H denotes entropy) value within a range that does not interfere with accuracy, equals $H(d \mid y)$. In other words, when $H(d \mid h)=H(d \mid y)$, i.e., the learned representation h contains as much domain information as the class labels, it does not affect the classification performance. After deriving the theorems, we propose a novel method adversarial feature learning with accuracy constraint (AFLAC), which leads to that invariance on adversarial training. Empirical validations show that the performance of AFLAC is superior to that of baseline methods, supporting the importance of considering domain-class dependency and the efficacy of the proposed approach for overcoming the issue.

在本文中，考虑到在权衡下优先考虑领域不变性可能会对领域泛化 (DG) 性能产生负面影响，我们提出在不干扰分类准确性的范围内最大化领域不变性 (图 1-(e))。我们首先扩展了关于领域对抗网络 (DAN) 的分析，该方法是一个广泛使用的隐式特征学习 (IFL) 方法，并推导出定理 1 和 2，展示了领域不变性对分类准确性产生负面影响的条件。在定理 3 中，我们展示了准确性约束下的领域不变性，我们将其定义为在不干扰准确性的范围内的最大值 $H(d \mid h)$ (H 表示熵)，等于 $H(d \mid y)$ 。换句话说，当 $H(d \mid h)=H(d \mid y)$ 时，即学习到的表示 h 包含与类别标签一样多的领域信息时，它不会影响分类性能。在推导出定理后，我们提出了一种新方法，即具有准确性约束的对抗特征学习 (AFLAC)，这导致了对抗训练中的不变性。实证验证表明，AFLAC 的性能优于基线方法，支持了考虑领域-类别依赖性的重要性以及所提方法克服该问题的有效性。

The main contributions of this paper can be summarized as follows. Firstly, we show that the implicit assumption of previous IFL methods, i.e., domain and class are statistically independent, is not valid in many real-world datasets, and it degrades the DG performance of them. Secondly, we theoretically show to what extent latent representation can become invariant to domains without interfering with classification accuracy. This is significant because the analysis guides the novel regularization approach that is suitable for our situation. Finally, we propose a novel method which improves domain invariance while maintaining classification performance, and it enjoys higher accuracy than the IFL methods on both synthetic and three real-world datasets.

本文的主要贡献可以总结如下。首先，我们表明之前的隐式特征学习方法的隐含假设，即领域和类别在统计上是独立的，在许多真实世界的数据集中并不成立，这降低了它们的领域泛化性能。其次，我们从理论上展示了潜在表示在多大程度上可以对领域变得不变，而不干扰分类准确性。这一点非常重要，因为该分析指导了适合我们情况的新正则化方法。最后，我们提出了一种新方法，该方法在保持分类性能的同时改善领域不变性，并且在合成数据集和三个真实世界数据集上比隐式特征学习方法具有更高的准确性。

2 Preliminary and Related Work

2 初步研究与相关工作

2.1 Problem Statement of Domain Generalization

2.1 领域泛化的问题陈述

Denote $\mathcal{X}, \mathcal{Y}$, and $\mathcal{D}$ as the input feature, class label, and domain spaces, respectively. With random variables $x \in \mathcal{X}, y \in \mathcal{Y}, d \in \mathcal{D}$ we can define the probability distribution for each domain as $p(x, y \mid d)$. For simplicity this paper assumes that y and d are discrete variables. In domain generalization, we are given a training dataset consisting of $\{x_i^s, y_i^s\}_{i=1}^{n^s}$ for all $s \in \{1,2,\ldots,m\}$, where each $\{x_i^s, y_i^s\}_{i=1}^{n^s}$ is drawn from the source domain $p(x, y \mid d=s)$. Using the training dataset, we train a classifier $g: \mathcal{X} \rightarrow \mathcal{Y}$, and use the classifier to predict labels of samples drawn from unknown target domain $p(x, y \mid d=t)$.

设定 $\mathcal{X}, \mathcal{Y}$ 、$\mathcal{D}$ 为输入特征、类别标签和领域空间，分别。通过随机变量 $x \in \mathcal{X}, y \in \mathcal{Y}, d \in \mathcal{D}$ ，我们可以将每个领域的概率分布定义为 $p(x, y \mid d)$ 。为了简便起见，本文假设 y 和 d 是离散变量。在领域泛化中，我们给定一个训练数据集，由所有 $s \in \{1,2, \ldots, m\}$ 的 $\{x_i^s, y_i^s\}_{i=1}^{n^s}$ 组成，其中每个 $\{x_i^s, y_i^s\}_{i=1}^{n^s}$ 是从源领域 $p(x, y \mid d=s)$ 中抽取的。利用训练数据集，我们训练一个分类器 $g: \mathcal{X} \rightarrow \mathcal{Y}$ ，并使用该分类器预测从未知目标领域 $p(x, y \mid d=t)$ 中抽取的样本的标签。

2.2 Related Work

2.2 相关工作

DG has been attracting considerable attention in recent years [27 28]. [18] showed that non-end-to-end DG methods such as DICA [27] and MTAE [11] do not tend to outperform vanilla CNN, thus end-to-end methods are desirable. End-to-end methods based on domain invariant representation can be divided into two categories: adversarial-learning-based methods such as DAN [9,33] and predefined-metric-based methods [10,20].

近年来，领域泛化 (DG) 引起了相当大的关注 [27 28]。[18] 显示，非端到端的 DG 方法，如 DICA [27] 和 MTAE [11]，往往不如普通的卷积神经网络 (CNN) 表现出色，因此端到端的方法是可取的。基于领域不变表示的端到端方法可以分为两类: 基于对抗学习的方法，如 DAN [9,33]，和基于预定义度量的方法 [10,20]。

In particular, our analysis and proposed method are based on DAN, which measures the invariance by using a domain classifier (also known as a discriminator) parameterized by deep neural networks and imposes regularization by deceiving it. Although DAN was originally invented for DA, [33] demonstrated its efficacy in DG. In addition, they intuitively explained the trade-off between classification accuracy and domain invariance, but did not suggest any solution to the problem except for carefully tuning a weighting parameter. AFLAC also relates to domain confusion loss [31] in that their encoders attempted to minimize Kullback-Leibler divergence (KLD) between the output distribution of the discriminators and some domain distribution ($p(d \mid y)$ in AFLAC and uniform distribution in [31]), rather than to deceive the discriminator as DAN.

特别是，我们的分析和提出的方法基于 DAN，它通过使用由深度神经网络参数化的领域分类器 (也称为判别器) 来测量不变性，并通过欺骗它来施加正则化。尽管 DAN 最初是为领域适应 (DA) 而发明的，[33] 证明了它在领域泛化 (DG) 中的有效性。此外，他们直观地解释了分类准确性与领域不变性之间的权衡，但除了仔细调整权重参数外，并没有提出任何解决方案。AFLAC 也与领域混淆损失 [31] 相关，因为它们的编码器试图最小化判别器输出分布与 AFLAC 中某个领域分布 ($p(d \mid y)$ 以及 [31] 中的均匀分布之间的 Kullback-Leibler 散度 (KLD)，而不是像 DAN 那样欺骗判别器。

Several studies that address DG without utilizing IFL have been conducted. For example, CCSA [26], CIDG [21], and CIDDG [22] proposed to make use of semantic alignment, which attempts to make latent representation given class label ($p(h \mid y)$) identical within source domains. This approach was originally proposed by [12] in the DA context, but its efficacy to overcome the trade-off problem is not obvious. Also, CIDDG, which is the only adversarial-learning-based semantic alignment method so far, needs the same number of domain classification networks as domains whereas ours needs only one. [37] also proposed a variant of adversarial-learning-based IFL method similar to ours, i.e., their method is also intended to maximize domain-invariance without affecting classification performance. Although their method needs to estimate true data distribution $p(y \mid x)$ with DNN, ours only needs to estimate $p(d \mid y)$, which is easily conducted when y and d are discrete random variable. CrossGrad [28], which is one of the recent state-of-the-art DG methods, utilizes data augmentation with adversarial examples. However, because the method relies on the assumption that y and d are independent, it might not be directly applicable to our setting. MLDG [19], MetaReg [2], and Feature-Critic [23], other state-of-the-art methods, are inspired by meta-learning. These methods make no assumption about the relation between y and d ; hence, they could be combined with our proposed method in principle.

已经进行了一些研究，探讨了不利用 IFL 的 DG。例如，CCSA [26]、CIDG [21] 和 CIDDG [22] 提出了利用语义对齐的方法，该方法试图使给定类别标签 ($p(h \mid y)$) 的潜在表示在源域内相同。该方法最初由 [12] 在 DA 背景下提出，但其克服权衡问题的有效性并不明显。此外，CIDDG 是迄今为止唯一基于对抗学习的语义对齐方法，它需要与域数相同数量的域分类网络，而我们的方法只需要一个。[37] 还提出了一种与我们相似的基于对抗学习的 IFL 方法，即他们的方法也旨在最大化域不变性而不影响分类性能。尽管他们的方法需要使用 DNN 估计真实数据分布 $p(y \mid x)$ ，但我们的方法只需估计 $p(d \mid y)$ ，当 y

和 d 是离散随机变量时，这一过程非常简单。CrossGrad [28] 是最近的最先进 DG 方法之一，利用对抗示例进行数据增强。然而，由于该方法依赖于 y 和 d 独立的假设，它可能不适用于我们的设置。MLDG [19]、MetaReg [2] 和 Feature-Critic [23] 等其他最先进的方法受到元学习的启发。这些方法对 y 和 d 之间的关系没有假设；因此，从原则上讲，它们可以与我们提出的方法结合使用。

As with our paper, 2132 also pointed out the importance of considering the types of distributional shifts that occur, and they address the shift of $p(y \mid x)$ across domains caused by the causal structure $y \rightarrow x$. However, the causal structure does not cause the trade-off problem as long as y and d are independent (Figure $1-(a, b)$) , thus it is essential to consider and address domain-class dependency problem. They also proposed to correct the domain-class dependency with the class prior-normalized weight, which enforces the prior probability for each class to be the same across domains. Its motivation is different from ours in that it is intended to avoid overfitting whereas we address the trade-off problem.

与我们的论文一样，2132 也指出了考虑发生的分布转变类型的重要性，他们讨论了因因果结构 $y \rightarrow x$ 导致的跨领域 $p(y \mid x)$ 的转变。然而，只要 y 和 d 是独立的，因果结构并不会导致权衡问题 (图 $1-(a, b)$))，因此考虑和解决领域-类别依赖问题是至关重要的。他们还提出通过类别先验归一化权重来修正领域-类别依赖性，这强制每个类别在各个领域的先验概率保持相同。其动机与我们的不同，旨在避免过拟合，而我们则关注权衡问题。

In DA,[36 12] address the situation where $p(y)$ changes across the source and target domains by correcting the change of $p(y)$ using unlabeled target domain data, which is often accomplished at the cost of classification accuracy for the source domain. However, this approach is not applicable (or necessary) to DG because we are agnostic on target domains and cannot run such adaptation step in DG. Instead, this paper is concerned with the change of $p(y)$ within source domain and proposes to maximize the classification accuracy for source domains while improving the domain invariance.

在领域适应 (DA) 中，[36 12] 解决了源领域和目标领域之间 $p(y)$ 变化的情况，通过使用未标记的目标领域数据来修正 $p(y)$ 的变化，这通常以牺牲源领域的分类准确性为代价。然而，这种方法不适用于 (或不必要) 领域泛化 (DG)，因为我们对目标领域是无偏见的，无法在 DG 中执行这样的适应步骤。相反，本文关注源领域内 $p(y)$ 的变化，并提出在提高领域不变性的同时最大化源领域的分类准确性。

It is worth mentioning that IFL has been used for many other context other than DG, e.g., DA [329], domain transfer [176], and fairness-aware classification [35 24 25]. However, adjusting it to each specific task is likely to improve performance. For example, in the fairness-aware classification task [25] proposed to optimize the fairness criterion directly instead of applying invariance to sensitive variables. By analogy, we adapted IFL for DG so as to address the domain-class dependency problem.

值得一提的是，IFL 已被用于许多其他上下文，而不仅仅是 DG，例如，领域适应 (DA)[329]、领域转移 [176] 和公平性感知分类 [35 24 25]。然而，将其调整为每个特定任务可能会提高性能。例如，在公平性感知分类任务中，[25] 提出了直接优化公平性标准，而不是对敏感变量应用不变性。类比而言，我们为 DG 调整了 IFL，以解决领域-类别依赖问题。

3 Our approach

3 我们的方法

3.1 Domain Adversarial Networks

3.1 领域对抗网络

In this section, we provide a brief overview of DAN [9], on which our analysis and proposed method are based. DAN trains a domain discriminator that attempts to predict domains from latent representation encoded by an encoder, while simultaneously training the encoder to remove domain information by deceiving the discriminator.

在本节中，我们提供了对 DAN [9] 的简要概述，我们的分析和提出的方法基于此。DAN 训练一个领域鉴别器，该鉴别器试图从编码器编码的潜在表示中预测领域，同时训练编码器通过欺骗鉴别器来去除领域信息。

Formally, we denote $f_E(x), q_M(y \mid h)$, and $q_D(d \mid h)(E, M$, and D are their parameters) as the deterministic encoder, probabilistic model of the label classifier, and that of domain discriminator, respectively. Then, the objective function of DAN is described as follows:

正式地，我们将 $f_E(x), q_M(y \mid h)$ 、$q_D(d \mid h)(E, M$ 和 D (它们的参数) 分别表示为确定性编码器、标签分类器的概率模型和领域鉴别器的概率模型。那么，DAN 的目标函数描述如下:

$$\min_{E,M}\max_{D} J\left(E,M,D\right)=\mathbb{E}_{p(x,d,y)}\left[-\gamma L_d+L_y\right], \tag{1}$$

$$\text{where } L_d \ := \ -\log q_D\left(d \mid h=f_E\left(x\right)\right), L_y \ := \ -\log q_M\left(y \mid h=f_E\left(x\right)\right) \ .$$

Here, the second term in Eq. 1 simply maximizes the log likelihood of q_M and f_E as well as in standard classification problems. On the other hand, the first term corresponds to a minimax game between the encoder and discriminator, where the discriminator $q_D\left(d \mid h\right)$ tries to predict d from h and the encoder $f_E\left(x\right)$ tries to fool $q_D\left(d \mid h\right)$.

在这里，公式 1 中的第二项简单地最大化 q_M 和 f_E 的对数似然，以及标准分类问题中的对数似然。另一方面，第一项对应于编码器和鉴别器之间的极小极大博弈，其中鉴别器 $q_D\left(d \mid h\right)$ 尝试从 h 中预测 d ，而编码器 $f_E\left(x\right)$ 尝试欺骗 $q_D\left(d \mid h\right)$ 。

As [33] originally showed, the minimax game ensures that the learned representation has no or little domain information, i.e., the representation becomes domain-invariant. This invariance ensures that the prediction from h to y is independent from d , and therefore hopefully facilitates the construction of a classifier capable of correctly handling samples drawn from unknown domains (Figure 1-(b)). Below is a brief explanation.

正如 [33] 最初所展示的，极小极大博弈确保学习到的表示没有或几乎没有领域信息，即表示变得领域不变。这种不变性确保从 h 到 y 的预测与 d 无关，因此希望能促进构建一个能够正确处理来自未知领域样本的分类器 (图 1-(b))。以下是简要说明。

Because h is a deterministic mapping of x , the joint probability distribution $p\left(h,d,y\right)$ can be defined as follows:

因为 h 是 x 的确定性映射，联合概率分布 $p\left(h,d,y\right)$ 可以定义如下:

$$\begin{aligned} p\left(h,d,y\right) &= \int_x p\left(x,d,h,y\right)dx = \int_x p\left(x,d,y\right)p\left(h \mid x\right)dx \\ &= \int_x p\left(x,d,y\right)\delta\left(f_E\left(x\right)=h\right)dx, \end{aligned} \tag{2}$$

and in the rest of the paper, we denote $p\left(h,d,y\right)$ as $\widetilde{p}_E\left(h,d,y\right)$ because it depends on the encoder’s parameter E . Using Eq. 2, Eq. 1 can be replaced as follows:

在本文的其余部分，我们将 $p\left(h,d,y\right)$ 表示为 $\widetilde{p}_E\left(h,d,y\right)$ ，因为它依赖于编码器的参数 E 。使用公式 2，公式 1 可以替换如下:

$$\min_{E,M}\max_{D} J\left(E,M,D\right)=\mathbb{E}_{\widetilde{p}_E(h,d,y)}\left[\gamma\log q_D\left(d \mid h\right)-\log q_M\left(y \mid h\right)\right]. \tag{3}$$

Assuming E is fixed, the solutions M^* and D^* to Eq. 3 satisfy $q_{M^*}\left(y \mid h\right)=\widetilde{p}_E\left(y \mid h\right)$ and $q_{D^*}\left(d \mid h\right)=\widetilde{p}_E\left(d \mid h\right)$. Substituting q_{M^*} and q_{D^*} into Eq. 3 enable us to obtain the following optimization problem depending only on E :

假设 E 是固定的，方程 3 的解 M^* 和 D^* 满足 $q_{M^*}\left(y \mid h\right)=\widetilde{p}_E\left(y \mid h\right)$ 和 $q_{D^*}\left(d \mid h\right)=\widetilde{p}_E\left(d \mid h\right)$ 。将 q_{M^*} 和 q_{D^*} 代入方程 3，使我们能够得到以下仅依赖于 E 的优化问题:

$$\min_{E} J\left(E\right)=-\gamma H_{\widetilde{p}_E}\left(d \mid h\right)+H_{\widetilde{p}_E}\left(y \mid h\right). \tag{4}$$

Solving Eq. 4 allows us to obtain the solutions M^*, D^* , and E^* , which are in Nash equilibrium. Here, $H_{\widetilde{p}_E}\left(d \mid h\right)$ means conditional entropy with the joint probability distribution $\widetilde{p}_E\left(d,h\right)$. Thus, minimizing the second term in Eq. 4 intuitively means learning (the mapping function f_E to) the latent representation h which contains as much information about y as possible. On the other hand, the first term can be regarded as a regularizer that attempts to learn h that is invariant to d .

求解方程 4 使我们能够获得解 M^*, D^* 和 E^* ，它们处于纳什均衡状态。这里，$H_{\widetilde{p}_E}\left(d \mid h\right)$ 表示具有联合概率分布 $\widetilde{p}_E\left(d,h\right)$ 的条件熵。因此，最小化方程 4 中的第二项直观上意味着学习 (映射函数 f_E 到) 潜在表示 h ，该表示尽可能包含关于 y 的信息。另一方面，第一项可以被视为一个正则化项，试图学习对 d 不变的 h 。

3.2 Trade-off Caused by Domain-Class Dependency

3.2 由领域-类别依赖引起的权衡

Here we show that the performance of DAN is impeded by the existence of domain-class dependency. Concretely, we show that the dependency causes the trade-off between classification accuracy and domain invariance: when d and y are statistically dependent, no values of E would be able to optimize the first and second term in Eq. 4 at the same time. Note that the following analysis also suggests that most IFL methods are negatively influenced by the dependency.

在这里，我们展示了 DAN 的性能受到领域-类别依赖的阻碍。具体而言，我们展示了这种依赖导致分类准确性与领域不变性之间的权衡: 当 d 和 y 在统计上相关时，没有任何 E 的值能够同时优化方程 4 中的第一项和第二项。请注意，以下分析还表明，大多数 IFL 方法受到这种依赖的负面影响。

To begin with, we consider only the first term in Eq. 4 and address the optimization problem:

首先，我们只考虑方程 4 中的第一项，并解决优化问题:

$$\min_E J_1(E) = -\gamma H_{\widetilde{p}_E}(d \mid h) \tag{5}$$

Using the property of entropy, $H_{\widetilde{p}_E}(d \mid h)$ is bounded:

利用熵的性质，$H_{\widetilde{p}_E}(d \mid h)$ 是有界的:

$$H_{\widetilde{p}_E}(d \mid h) \leq H(d) \tag{6}$$

Thus, Eq. 5 has the solution E_1^* which satisfies the following condition:

因此，方程 5 的解 E_1^* 满足以下条件:

$$H_{\widetilde{p}_{E_1^*}}(d \mid h) = H(d) \tag{7}$$

Eq. 7 suggests that the regularizer in DAN is intended to remove all information about domains from latent representation h , thereby ensuring the independence of domains and latent representation.

方程 7 表明 DAN 中的正则化器旨在从潜在表示 h 中移除所有关于领域的信息，从而确保领域与潜在表示的独立性。

Next, we consider only the second term in Eq. 4, thereby addressing the following optimization problem:

接下来，我们仅考虑方程 4 中的第二项，从而解决以下优化问题:

$$\min_E J_2(E) = H_{\widetilde{p}_E}(y \mid h) \tag{8}$$

Considering h is the mapping of x , i.e., $h = f_E(x)$, the solution E_2^* to Eq. 8 satisfies the following equation:

考虑 h 是 x 的映射，即 $h = f_E(x)$ ，方程 8 的解 E_2^* 满足以下方程:

$$H_{\widetilde{p}_{E_2^*}}(y \mid h) = H(y \mid x). \tag{9}$$

Here we obtain E_1^* and E_2^* , which can achieve perfect invariance and optimal classification accuracy, respectively. Using them, we can obtain the following theorem, which shows the existence of the trade-off between invariance and accuracy: perfect invariance (E_1^*) and optimal classification accuracy (E_2^*) cannot be achieved at the same time.

在这里，我们获得 E_1^* 和 E_2^* ，它们分别可以实现完美不变性和最佳分类准确性。利用它们，我们可以得到以下定理，表明不变性与准确性之间存在权衡: 完美不变性 (E_1^*) 和最佳分类准确性 (E_2^*) 不能同时实现。

Theorem 1 When $H(y \mid x) = 0$, i.e., there is no labeling error, and $H(d) > H(d \mid y)$, i.e., the domain and class are statistically dependent, $E_1^* \neq E_2^*$ holds.

定理 1 当 $H(y \mid x) = 0$ ，即没有标记错误，并且 $H(d) > H(d \mid y)$ ，即领域和类别在统计上是相关的，$E_1^* \neq E_2^*$ 成立。

Proof 1 Assume $E_1^* = E_2^* = E^*$. Using the properties of entropy, we can obtain the following:

证明 1 假设 $E_1^* = E_2^* = E^*$ 。利用熵的性质，我们可以得到以下结果:

$$H_{\widetilde{p}_E}(d \mid h) \leq H_{\widetilde{p}_E}(d, y \mid h) = H_{\widetilde{p}_E}(d \mid h, y) + H_{\widetilde{p}_E}(y \mid h) \leq H_{\widetilde{p}_E}(d \mid y) + H_{\widetilde{p}_E}(y \mid h).$$

(10)

Substituting $H_{\widetilde{p}_{E^*}}(y \mid h) = H(y \mid x)$ and $H_{\widetilde{p}_{E^*}}(d \mid h) = H(d)$ into Eq. 10, we can obtain the following condition:

将 $H_{\widetilde{p}_{E^*}}(y \mid h) = H(y \mid x)$ 和 $H_{\widetilde{p}_{E^*}}(d \mid h) = H(d)$ 代入方程 10，我们可以得到以下条件:

$$H(d) - H(d \mid y) \leq H(y \mid x). \tag{11}$$

Because the domain and class are dependent on each other, the following condition holds:

由于领域和类别相互依赖，以下条件成立:

$$0 < H(d) - H(d \mid y) \leq H(y \mid x), \tag{12}$$

but Eq. 12 contradicts with $H(y \mid x) = 0$. Thus, $E_1^* \neq E_2^*$.

但方程 12 与 $H(y \mid x) = 0$ 矛盾。因此，$E_1^* \neq E_2^*$ 。

Theorem 1 shows that the domain-class dependency causes the trade-off problem. Although it assumes $H(y \mid x) = 0$ for simplicity, we cannot know the true value of $H(y \mid x)$ and there are many cases in which little or no labeling errors occur and thus $H(y \mid x)$ is close to 0 .

定理 1 表明领域-类别依赖性导致了权衡问题。尽管为了简化假设了 $H(y \mid x) = 0$ ，我们无法知道 $H(y \mid x)$ 的真实值，并且在许多情况下几乎没有标记错误发生，因此 $H(y \mid x)$ 接近 0。

In addition, we can omit the assumption and obtain a more general result:

此外，我们可以省略假设并获得更一般的结果:

Theorem 2 When $I(d; y) := H(d) - H(d \mid y) > H(y \mid x), E_1^* \neq E_2^*$ holds.

定理 2 当 $I(d; y) := H(d) - H(d \mid y) > H(y \mid x), E_1^* \neq E_2^*$ 成立。

Proof 2 Similar to Proof 1, we assume that $E_1^* = E_2^*$ and thus Eq. 11 is obtained. Obviously, Eq. 11 does not hold when $H(d) - H(d \mid y) > H(y \mid x)$.

证明 2 类似于证明 1，我们假设 $E_1^* = E_2^*$ ，因此得到了方程 11。显然，当 $H(d) - H(d \mid y) > H(y \mid x)$ 时，方程 11 不成立。

Theorem 2 shows that when the mutual information of the domain and class $I(d; y)$ is greater than the labeling error $H(y \mid x)$, the trade-off between invariance and accuracy occurs. Then, although we cannot know the true value of $H(y \mid x)$, the performance of DAN and other IFL methods are likely to decrease when $I(d; y)$ has large value.

定理 2 表明，当领域与类别的互信息 $I(d; y)$ 大于标记错误 $H(y \mid x)$ 时，出现了不变性与准确性之间的权衡。因此，尽管我们无法知道 $H(y \mid x)$ 的真实值，但当 $I(d; y)$ 的值较大时，DAN 和其他 IFL 方法的性能可能会下降。

3.3 Accuracy-Constrained Domain Invariance

3.3 准确性约束下的领域不变性

If we cannot avoid the trade-off, the next question is to decide how to accommodate it, i.e., to what extent the representation should become domain-invariant for DG tasks. Here we provide the notion of accuracy-constrained domain invariance, which is the maximum domain invariance within a range that does not interfere with the classification accuracy. The reason for the constraint is that the primary purpose of DG is the classification for unseen domains rather than the invariance itself, and the improvement of the invariance could detrimentally affect the performance. For example, in WISDM, if we know the target activity was performed by a young rather than an old man, we might predict the activity to be jogging with a higher probability; thus, we would have to avoid removing such domain information that may be useful in the classification task.

如果我们无法避免权衡，接下来的问题是决定如何适应它，即表示应在多大程度上变得对 DG 任务不变。在这里，我们提供了准确性约束下的领域不变性的概念，这是在不干扰分类准确性的范围内的最大领域不变性。约束的原因在于，DG 的主要目的是对未见领域进行分类，而不是不变性本身，改善不变性可能会对性能产生不利影响。例如，在 WISDM 中，如果我们知道目标活动是由年轻人而不是老年人执行的，我们可能会以更高的概率预测该活动为慢跑；因此，我们必须避免去除可能对分类任务有用的领域信息。

Theorem 3 Define accuracy-constrained domain invariance as the maximum $H_{\widetilde{p}_E}(d \mid h)$ value under the constraint that $H(y \mid x) = 0$, i.e., there is no labeling error, and classification accuracy is maximized, i.e., $H_{\widetilde{p}_E}(y \mid h) = H(y \mid x)$. Then, accuracy-constrained domain invariance equals $H(d \mid y)$.

定理 3 将准确性约束下的领域不变性定义为在约束条件 $H(y \mid x) = 0$ 下的最大 $H_{\widetilde{p}_E}(d \mid h)$ 值，即没有标记错误，并且分类准确性最大化，即 $H_{\widetilde{p}_E}(y \mid h) = H(y \mid x)$ 。然后，准确性约束下的领域不变性等于 $H(d \mid y)$ 。

Proof 3 Using Eq. 10 and $H_{\widetilde{p}_E}(y \mid h) = H(y \mid x)$, the following inequation holds:
证明 3 使用方程 10 和 $H_{\widetilde{p}_E}(y \mid h) = H(y \mid x)$ ，以下不等式成立:

$$H_{\widetilde{p}_E}(d \mid h) \leq H(y \mid x) + H(d \mid y) . \tag{13}$$

Substituting $H(y \mid x) = 0$ into Eq. 13, the following inequation holds:
将 $H(y \mid x) = 0$ 代入方程 13，以下不等式成立:

$$H_{\widetilde{p}_E}(d \mid h) \leq H(d \mid y) . \tag{14}$$

Thus, the maximum $H_{\widetilde{p}_E}(d \mid h)$ value under the optimal classification accuracy constraint is $H(d \mid y)$.

因此，在最佳分类准确性约束下，最大 $H_{\widetilde{p}_E}(d \mid h)$ 值为 $H(d \mid y)$ 。

Note that we could improve the invariance more when $H(y \mid x) > 0$ (that is obvious considering Eq. 13), but we cannot know the true value of $H(y \mid x)$ as we discussed in Sec. 3.2. Thus, accuracy-constrained domain invariance can be viewed as the worst-case gurantee.

注意，当 $H(y \mid x) > 0$ 时，我们可以进一步提高不变性 (考虑方程 13 这一点显而易见)，但正如我们在第 3.2 节中讨论的那样，我们无法知道 $H(y \mid x)$ 的真实值。因此，准确性约束下的领域不变性可以被视为最坏情况的保证。

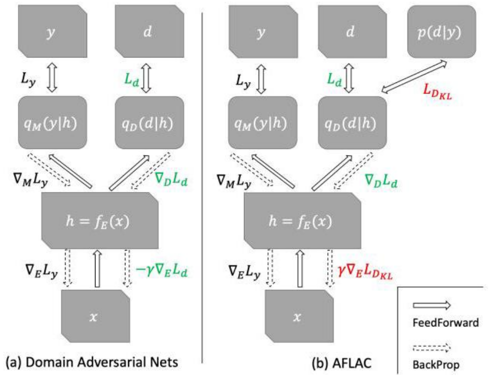


Fig. 2. Comparative illustration of DAN and AFLAC. (a) The classifier and discriminator try to minimize L_y and L_d , respectively. The encoder tries to minimize L_y and maximize L_d (fool the discriminator). (b) The discriminator tries to approximate true $\widetilde{p}_E(d \mid h)$ by minimizing L_d . The encoder tries to minimize divergence between $\widetilde{p}_E(d \mid h)$ and $p(d \mid y)$ by minimizing $L_{D_{KL}}$.

图 2. DAN 和 AFLAC 的比较示意图。(a) 分类器和鉴别器分别尝试最小化 L_y 和 L_d 。编码器尝试最小化 L_y 并最大化 L_d (欺骗鉴别器)。(b) 鉴别器通过最小化 L_d 尝试逼近真实的 $\widetilde{p}_E(d \mid h)$ 。编码器通过最小化 $L_{D_{KL}}$ 尝试最小化 $\widetilde{p}_E(d \mid h)$ 和 $p(d \mid y)$ 之间的散度。

3.4 Proposed Method

3.4 提出的方法

Based on the above analysis, the remaining challenge is to determine how to achieve accuracy-constrained domain invariance, i.e., imposing regularization such that makes $H_{\widetilde{p}_E}(d \mid h) = H(d \mid y)$ holds. Although DAN might be able to achieve this condition by carefully tuning the strength of the regularizer (γ in Eq. 1) , such tuning is time-consuming and impractical, as suggested by our experiments. Alternatively, we propose a novel method named AFLAC by modifying the regularization term of DAN: whereas the encoder of DAN attempts to fool the discriminator, that of AFLAC attempts to directly minimize the KLD between $p(d \mid y)$ and $q_D(d \mid h)$. Formally, AFLAC solves the following joint optimization problem by alternating gradient descent.

基于上述分析，剩下的挑战是确定如何实现准确性约束下的领域不变性，即施加正则化以使 $H_{\widetilde{p}_E}(d \mid h) = H(d \mid y)$ 成立。尽管 DAN 可能能够通过仔细调整正则化器 (γ in Eq. 1) 的强度来实现这一条件，但正如我们的实验所示，这种调整耗时且不切实际。作为替代，我们提出了一种名为 AFLAC 的新方法，通过修改 DAN 的正则化项: 而 DAN 的编码器试图欺骗鉴别器，AFLAC 的编码器

则试图直接最小化 $p\left(d\mid y\right)$ 和 $q_D\left(d\mid h\right)$ 之间的 KLD。正式地，AFLAC 通过交替梯度下降解决以下联合优化问题。

$$\min_D W\left(E,D\right)=\mathbb{E}_{p(x,d)}\left[L_d\right] \tag{15}$$

$$\min_{E,M} V\left(E,M\right)=\mathbb{E}_{p(x,d,y)}\left[\gamma L_{D_{KL}}+L_y\right], \tag{16}$$

$$\text{where } L_{D_{KL}} \;:=\; D_{KL}\left[p\left(d\mid y\right)\mid q_D\left(d\mid h=f_E\left(x\right)\right)\right]\ .$$

The minimization of L_y and L_d , respectively, means maximization of the log-likelihood of q_M and q_D as well as in DAN. However, the minimization of $L_{D_{KL}}$ differs from the regularizer of DAN in that it is intended to satisfy $q_D\left(d\mid h\right)=p\left(d\mid y\right)$. And if $q_D\left(d\mid h\right)$ well approximates $\widetilde{p}_E\left(d\mid h\right)$ by the minimization of L_d in Eq. 15, the minimization of $L_{D_{KL}}$ leads to $\widetilde{p}_E\left(d\mid h\right)=p\left(d\mid y\right)$. Figure 2-(b) outlines the training of AFLAC.

对 L_y 和 L_d 的最小化分别意味着 q_M 和 q_D 的对数似然最大化，以及在 DAN 中的最大化。然而，$L_{D_{KL}}$ 的最小化与 DAN 的正则化器不同，因为它旨在满足 $q_D\left(d\mid h\right)=p\left(d\mid y\right)$ 。如果 $q_D\left(d\mid h\right)$ 通过在方程 15 中最小化 L_d 来很好地逼近 $\widetilde{p}_E\left(d\mid h\right)$ ，那么 $L_{D_{KL}}$ 的最小化将导致 $\widetilde{p}_E\left(d\mid h\right)=p\left(d\mid y\right)$ 。图 2-(b) 概述了 AFLAC 的训练过程。

Here we formally show that AFLAC is intended to achieve $H_{\widetilde{p}_E}\left(d\mid h\right)=H\left(d\mid y\right)$ (accuracy-constrained domain invariance) by a Nash equilibrium analysis smilar to [13,33]. As well as in Section 3.1, D^* and M^* , which are the solutions to Eqs. 15,16 with fixed E , satisfy $q_D^*=\widetilde{p}_E\left(d\mid h\right)$ and $q_M^*=\widetilde{p}_E\left(y\mid h\right)$, respectively. Thus, $\bar{V}$ in Eq. 16 can be written as follows:

在这里，我们正式证明 AFLAC 旨在通过类似于 [13,33] 的纳什均衡分析来实现 $H_{\widetilde{p}_E}\left(d\mid h\right)=H\left(d\mid y\right)$ (精度约束的领域不变性)。如第 3.1 节所述，D^* 和 M^* 是在固定 E 的情况下方程 15 和 16 的解，分别满足 $q_D^*=\widetilde{p}_E\left(d\mid h\right)$ 和 $q_M^*=\widetilde{p}_E\left(y\mid h\right)$ 。因此，方程 16 中的 $\bar{V}$ 可以写成如下形式:

$$V\left(E\right)=\mathbb{E}\left[\gamma D_{KL}\left[p\left(d\mid y\right)\mid \widetilde{p}_E\left(d\mid h\right)\right]\right]+H_{\widetilde{p}_E}\left(y\mid h\right). \tag{17}$$

E^* , which is the solution to Eq. 17 and in Nash equilibrium, satisfies not only $H_{\widetilde{p}_{E^*}}\left(y\mid h\right)=H\left(y\mid x\right)$ (optimal classification accuracy) but also

E^* 是方程 17 的解，并且在纳什均衡中，不仅满足 $H_{\widetilde{p}_{E^*}}\left(y\mid h\right)=H\left(y\mid x\right)$ (最佳分类精度)，而且还满足

$\mathbb{E}_{h,y\sim\widetilde{p}_{E^*}(h,y)}\left[D_{KL}\left[p\left(d\mid y\right)\mid \widetilde{p}_{E^*}\left(d\mid h\right)\right]\right]=0$, which is a sufficient condition for $H_{\widetilde{p}_{E^*}}\left(d\mid h\right)=H\left(d\mid y\right)$ by the definition of the conditional entropy.

$\mathbb{E}_{h,y\sim\widetilde{p}_{E^*}(h,y)}\left[D_{KL}\left[p\left(d\mid y\right)\mid \widetilde{p}_{E^*}\left(d\mid h\right)\right]\right]=0$ 是根据条件熵定义的 $H_{\widetilde{p}_{E^*}}\left(d\mid h\right)=H\left(d\mid y\right)$ 的充分条件。

In training, $p\left(x,d,y\right)$ in the objectives (Eqs. 15,16) is approximated by empirical distribution composed of the training data obtained from m source domains, i.e., $\left\{x_i^{(1)},y_i^{(1)},d=1\right\}_{i=1}^{n^{(1)}},\dots,\left\{x_i^{(m)},y_i^{(m)},d=m\right\}_{i=1}^{n^{(m)}}$. Also, $p\left(d\mid y\right)$ used in Eq. 16 can be replaced by the maximum likelihood or maximum a posteriori estimator of it. Note that, we could use some distances other than $D_{KL}\left[p(d|y)|\, q_D\left(d\mid h\right)\right]$ in Eq. 16, e.g., $D_{KL}\left[q_D\left(d\mid h\right)\mid p\left(d\mid y\right)\right]$, but in doing so, we could not observe performance gain, hence we discontinued testing them.

在训练中，$p\left(x,d,y\right)$ 在目标中 (方程 15,16) 通过由来自 m 源域获得的训练数据组成的经验分布来近似，即 $\left\{x_i^{(1)},y_i^{(1)},d=1\right\}_{i=1}^{n^{(1)}},\dots,\left\{x_i^{(m)},y_i^{(m)},d=m\right\}_{i=1}^{n^{(m)}}$ 。此外，方程 16 中使用的 $p\left(d\mid y\right)$ 可以被其最大似然或最大后验估计量替代。请注意，我们可以在方程 16 中使用一些其他距离，而不是 $D_{KL}\left[p(d|y)|\, q_D\left(d\mid h\right)\right]$ ，例如 $D_{KL}\left[q_D\left(d\mid h\right)\mid p\left(d\mid y\right)\right]$ ，但这样做时，我们没有观察到性能提升，因此我们停止了对它们的测试。

4 Experiments

4 实验

4.1 Datasets

4.1 数据集

Here we provide a brief overview of one synthetic and three real-world datasets (PACS, WISDM, IEMOCAP) used for the performance evaluation. Although WISDM and IEMOCAP have not been widely used in DG studies, previous human activity recognition and speech emotion recognition studies (e.g., [185]) used them in the domain generalization setting (i.e., source and target domains are disjoint), so they can be regarded as the practical use case of domain generalization. The concrete sample sizes for each d and y , and the network architectures for each dataset are shown in supplementary 1

在这里，我们提供一个合成数据集和三个真实世界数据集 (PACS, WISDM, IEMOCAP) 的简要概述，这些数据集用于性能评估。尽管 WISDM 和 IEMOCAP 在领域泛化研究中尚未被广泛使用，但之前的人类活动识别和语音情感识别研究 (例如，[185]) 在领域泛化设置中使用了它们 (即源域和目标域是不相交的)，因此它们可以被视为领域泛化的实际应用案例。每个 d 和 y 的具体样本大小，以及每个数据集的网络架构在补充材料 1 中显示。

BMNISTR We created the Biased and Rotated MNIST dataset (BMNISTR) by modifying the sample size of the popular benchmark dataset MNISTR [11, such that the class distribution differed among the domains. In MNISTR, each class is represented by 10 digits. Each domain was created by rotating images by 15 degree increments:0,15,30,45,60, and 75 (referred to as M0,..., M75). Each image was cropped to 16×16 in accordance with [11]. We created three variants of MNISTR that have different types of domain-class dependency, referred to as BMNISTR-1 through BMNISTR-3. As shown in Table 1-left, BMNISTR-1, -2 have similar trends but different degrees of dependency, whereas BMNISTR-1 and BMNISTR-3 differ in terms of their trends.

BMNISTR 我们通过修改流行基准数据集 MNISTR 的样本大小创建了偏置和旋转 MNIST 数据集 (BMNISTR) [11]，使得类分布在不同领域之间存在差异。在 MNISTR 中，每个类别由 10 个数字表示。每个领域是通过将图像旋转 15 度增量创建的:0、15、30、45、60 和 75(称为 M0，...，M75)。每个图像根据 [11] 被裁剪为 16×16 。我们创建了三种不同类型的领域-类别依赖关系的 MNISTR 变体，称为 BMNISTR-1 到 BMNISTR-3。如表 1 左侧所示，BMNISTR-1 和 BMNISTR-2 具有相似的趋势，但依赖程度不同，而 BMNISTR-1 和 BMNISTR-3 在趋势上存在差异。

PACS The PACS dataset [18] contains 9991 images across 7 categories (dog, elephant, giraffe, guitar, house, horse, and person) and 4 domains comprising different stylistic depictions (Photo, Art painting, Cartoon, and Sketch). It has domain-class dependency probably owing to the data characteristics. For example, $p(y=\text{ person } \mid d=\text{ Phot })$ is much higher than $p(y=\text{ person } \mid d=\text{ Sketch })$, indicating that photos of a person are easier to obtain than those of animals, but sketches of persons are more difficult to obtain than those of animals in the wild. For training, we used the ImageNet pre-trained AlexNet CNN [15] as the base network, following previous studies [18 19]. The two-FC-layer discriminator was connected to the last FC layer, following [9].

PACS PACS 数据集 [18] 包含 9991 张图像，分为 7 类 (狗、大象、长颈鹿、吉他、房屋、马和人) 以及 4 个领域，涵盖不同的风格表现 (照片、艺术画、卡通和素描)。由于数据特征，它可能具有领域-类别依赖性。例如，$p(y=\text{ person } \mid d=\text{ Phot })$ 远高于 $p(y=\text{ person } \mid d=\text{ Sketch })$ ，这表明获取人的照片比获取动物的照片更容易，但获取人的素描比获取野生动物的素描更困难。为了训练，我们使用了 ImageNet 预训练的 AlexNet CNN [15] 作为基础网络，遵循之前的研究 [18 19]。两个全连接层的鉴别器连接到最后一个全连接层，遵循 [9]。

WISDM The WISDM Activity Prediction dataset contains sensor data of accelerometers of six human activities (walking, jogging, upstairs, downstairs, sitting, and standing) performed by 36 users (domains). WISDM has the dependency for the reason noted in Sec. 1 . In data preprocessing, we use the sliding-window procedure with 60 frames (= 3 seconds) referring to [1], and the total number of samples was 18210. We parameterized the encoder using three 1-D convolution layers followed by one FC layer and the classifier by logistic regression, following previous studies [34,14].

[1] Code and Supplementary are available at https://github.com/akuzeee/AFLAC

[1] 代码和补充材料可在 https://github.com/akuzeee/AFLAC 获取

WISDM WISDM 活动预测数据集包含来自 36 个用户 (领域) 进行的六种人类活动 (步行、慢跑、上楼、下楼、坐着和站着) 的加速度计传感器数据。WISDM 具有在第 1 节中提到的依赖关系。在数据预处理过程中，我们使用滑动窗口程序，窗口大小为 60 帧 (= 3 seconds)，参考文献 [1]，总样本数为 18210。我们使用三个 1-D 卷积层和一个全连接层对编码器进行参数化，并通过逻辑回归对分类器进行参数化，遵循之前的研究 [34,14]。

IEMOCAP The IEMOCAP dataset [4] is the popular benchmark dataset for speech emotion recognition (SER), which aims at recognizing the correct emotional state of the speaker from speech signals. It contains a total of 10039 utterances pronounced by ten actors (domains, referred to as Ses01F, Ses01M through Ses05F, Ses05M) with emotional categories, and we only consider the four emotional categories (happy, angry, sad, and neutral) referring to [5.8]. Also, we refered to [5] about data preprocessing: we split the speech signal into equal-length segments of 3s, and extracted 40-dimensional log Mel-spectrogram, its deltas, and delta-deltas. We parameterized the encoder using three 2-D convolution layers followed by one FC layer and the classifier by logistic regression.

IEMOCAP IEMOCAP 数据集 [4] 是语音情感识别 (SER) 的热门基准数据集，旨在从语音信号中识别说话者的正确情感状态。它包含由十位演员 (领域，称为 Ses01F、Ses01M 到 Ses05F、Ses05M) 发音的总共 10039 个话语，具有情感类别，我们仅考虑四个情感类别 (快乐、愤怒、悲伤和中性)，参考文献 [5.8]。此外，我们参考了 [5] 关于数据预处理: 我们将语音信号分割为等长的 3 秒段，并提取 40 维的对数 Mel 频谱图、其增量和增量的增量。我们使用三个 2-D 卷积层和一个全连接层对编码器进行参数化，并通过逻辑回归对分类器进行参数化。

4.2 Baselines

4.2 基线

To demonstrate the efficacy of the proposed method AFLAC, we compared it with vanilla CNN and adversarial-learning-based methods. Specifically, (1) CNN is a vanilla convolutional networks trained on the aggregation of data from all source domains. Although CNN has no special treatments for DG, [18] reported that it outperforms many traditional DG methods. (2) DAN [33] is expected to generalize across domains utilizing domain-invariant representation, but it can be affected by the trade-off between domain invariance and accuracy as explained in Section 3.2. (3) CIDDG is our re-implementation of the method proposed in [22], which is designed to achieve semantic alignment on adversarial training. Additionally, we used (4) AFLAC-Abl, which is a version of AFLAC modified for ablation studies. AFLAC-Abl replaces $D_{KL}\left[p(d|y)|\, q_D\left(d \mid h\right)\right]$ in Eq. 16 of $D_{KL}\left[p\left(d\right) \mid q_D\left(d \mid h\right)\right]$, thus it attempts to learn the representation that is perfectly invariant to domains or make $H\left(d \mid h\right) = H\left(d\right)$ hold as well as DAN. Comparing AFLAC and AFLAC-Abl, we measured the genuine effect of taking domain-class dependency into account. When training AFLAC and AFLAC-Abl, we cannot obtain true $p\left(d \mid y\right)$ and $p\left(d\right)$, hence we used their maximum likelihood estimators for calculating the KLD terms.

为了证明所提出的方法 AFLAC 的有效性，我们将其与普通 CNN 和基于对抗学习的方法进行了比较。具体而言，(1) CNN 是一种普通卷积网络，训练时使用来自所有源域的数据聚合。尽管 CNN 对于领域泛化没有特殊处理，[18] 报告称它在许多传统领域泛化方法中表现优于。(2) DAN [33] 预计能够利用领域不变表示在不同领域之间进行泛化，但它可能受到领域不变性与准确性之间权衡的影响，如第 3.2 节所解释的那样。(3) CIDDG 是我们对 [22] 中提出的方法的重新实现，旨在实现对抗训练中的语义对齐。此外，我们使用了 (4) AFLAC-Abl，这是一个为消融研究修改的 AFLAC 版本。AFLAC-Abl 在 $D_{KL}\left[p\left(d\right) \mid q_D\left(d \mid h\right)\right]$ 的公式 16 中替换了 $D_{KL}\left[p(d|y)|\, q_D\left(d \mid h\right)\right]$ ，因此它试图学习对领域完全不变的表示，或者使 $H\left(d \mid h\right) = H\left(d\right)$ 与 DAN 一样成立。通过比较 AFLAC 和 AFLAC-Abl，我们测量了考虑领域类别依赖性的真实效果。在训练 AFLAC 和 AFLAC-Abl 时，我们无法获得真实的 $p\left(d \mid y\right)$ 和 $p\left(d\right)$ ，因此我们使用它们的最大似然估计量来计算 KLD 项。

Table 1. Left: Sample sizes for each domain-class pair in BMNISTR. Those for the classes $0 \sim 4$ are variable across domains, whereas the classes $5 \sim 9$ have identical sample sizes across domains. Right: Mean F-measures for the classes $0 \sim 4$ and classes $5 \sim 9$ with the target domain M0. RI denotes relative improvement of AFLAC to AFLAC-Abl

表 1. 左侧:BMNISTR 中每个领域类别对的样本大小。类别 $0 \sim 4$ 的样本大小在不同领域之间是可变的，而类别 $5 \sim 9$ 在不同领域之间具有相同的样本大小。右侧: 目标领域 M0 中类别 $0 \sim 4$ 和类别 $5 \sim 9$ 的平均 F 测量值。RI 表示 AFLAC 相对于 AFLAC-Abl 的相对改进。

Dataset	Class	M0	M15	M30	M45	M60	M75
BMNISTR-1	04	100	85	70	55	40	25
	59	100	100	100	100	100	100
BMNISTR-2	$0\sim4$	100	90	80	70	60	50
	59	100	100	100	100	100	100
BMNISTR-3	04	100	25	100	25	100	25
	$5\sim9$	100	100	100	100	100	100

数据集	类别	M0	M15	M30	M45	M60	M75
BMNISTR-1	04	100	85	70	55	40	25
	59	100	100	100	100	100	100
BMNISTR-2	$0\sim4$	100	90	80	70	60	50
	59	100	100	100	100	100	100
BMNISTR-3	04	100	25	100	25	100	25
	$5\sim9$	100	100	100	100	100	100

Dataset	Class	CNN	DAN	CIDDG	AFLAC -Abl	AFLAC	RI
BMNISTR-1	$0\sim4$	83.86	84.54	87.50	87.46	90.62	3.6%
	59	83.90	85.24	87.46	86.46	88.10	1.9%
BMNISTR-2	$0\sim4$	82.54	85.30	87.64	88.60	89.64	1.2%
	59	82.18	85.80	86.74	87.60	89.04	1.6%
BMNISTR-3	04	71.26	79.22	76.76	76.56	80.02	4.5%
	$5\sim9$	78.62	83.14	82.64	82.94	82.80	-0.2%

数据集	类别	CNN	DAN	CIDDG	AFLAC -Abl	AFLAC	RI
BMNISTR-1	$0\sim4$	83.86	84.54	87.50	87.46	90.62	3.6%
	59	83.90	85.24	87.46	86.46	88.10	1.9%
BMNISTR-2	$0\sim4$	82.54	85.30	87.64	88.60	89.64	1.2%
	59	82.18	85.80	86.74	87.60	89.04	1.6%
BMNISTR-3	04	71.26	79.22	76.76	76.56	80.02	4.5%
	$5\sim9$	78.62	83.14	82.64	82.94	82.80	-0.2%

4.3 Experimental Settings

4.3 实验设置

For all the datasets and methods, we used RMSprop for optimization. Further, we set the learning rate, batch size, and the number of iterations as $5e-4,128$, and 10k for BMNISTR; 5e-5,64, and 10k for PACS; 1e-4,64, and 10k for IEMO-CAP; 5e-4 (with exponential decay with decay step 18k and 24k , and decay rate 0.1),128, and 30k for WISDM, respectively. Also, we used the annealing of weighting parameter γ proposed in [9], and unless otherwise mentioned chose γ from $\{0.0001, 0.001, 0.01, 0.1, 1, 10\}$ for DAN, CIDDG, AFLAC-Abl, and AFLAC. Specifically, on BMNISTR and PACS, we employed a leave-one-domain-out setting [11], i.e., we chose one domain as target and used the remaining domains as source data. Then we split the source data into 80% of training data and 20% of validation data, assuming that target data are not absolutely available in the training phase. On IEMOCAP, we chose the best γ from

对于所有数据集和方法，我们使用 RMSprop 进行优化。此外，我们将学习率、批量大小和迭代次数设置为 $5e-4,128$ 和 10k 用于 BMNISTR；5e-5、64 和 10k 用于 PACS；1e-4、64 和 10k 用于 IEMO-CAP；5e-4(具有衰减步长 18k 和 24k 的指数衰减，衰减率为 0.1)、128 和 30k 用于 WISDM。此外，我们使用了 [9] 中提出的权重参数 γ 的退火，除非另有说明，否则选择 γ 来自 $\{0.0001, 0.001, 0.01, 0.1, 1, 10\}$ 用于 DAN、CIDDG、AFLAC-Abl 和 AFLAC。具体而言，在 BMNISTR 和 PACS 上，我们采用了留一领域法设置 [11]，即我们选择一个领域作为目标，使用其余领域作为源数据。然后，我们将源数据拆分为 80% 的训练数据和 20% 的验证数据，假设目标数据在训练阶段并不绝对可用。在 IEMOCAP 上，我们选择了最佳的 γ 。

$\{0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000\}$ using disjoint validation domain, referring to [8.5]. On WISDM, we randomly selected $< 20/16 >$ users as $<$ source / target$>$ domains, and split the source data into training and validation data because one-domain-leave-out evaluation is computationally expensive. Then, we conducted experiments multiple times with different random weight initialization; we trained the models on 10,20, and 20 seeds in BMNISTR, WISDM, and IEMOCAP, chose the best hyperparameter that achieved the highest validation accuracies measured in each epoch, and reported the mean scores (accuracies and F-measures) for the hyperparameter. On PACS, because it requires a long time to train on, we chose the best γ from $\{0.0001, 0.001, 0.01, 0.1\}$ after three experiments, and reported the mean scores in experiments with 15 seeds.

$\{0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000\}$ 使用不相交的验证域，参考文献 [8.5]。在 WISDM 上，我们随机选择了 < 20/16 > 用户作为 < 源/目标域，并将源数据分割为训练数据和验证数据，因为单域留出评估的计算成本较高。然后，我们进行了多次实验，使用不同的随机权重初始化；我们在 BMNISTR、WISDM 和 IEMOCAP 上分别使用 10、20 和 20 个种子训练模型，选择在每个时期测量的最高验证准确率所对应的最佳超参数，并报告该超参数的平均得分 (准确率和 F 值)。在 PACS 上，由于训练时间较长，我们在三次实验后从 $\{0.0001, 0.001, 0.01, 0.1\}$ 中选择了最佳 γ ，并报告了使用 15 个种子的实验中的平均得分。

4.4 Results

4.4 结果

We first investigated the extent to which domain-class dependency affects the performance of the IFL methods. In Table 1-right, we compared the mean F-measures for the classes 0 through 4 and classes 5 through 9 in BMNISTR with the target domain M0. Recall that the sample sizes for the classes $0 \sim 4$ are variable across domains, whereas the classes $5 \sim 9$ have identical sample sizes across domains (Table 1-left). The F-measures show that AFLAC outperformed baselines in most dataset-class pairs, which supports that domain-class dependency reduces the performance of domain-invariance-based methods and that AFLAC can mitigate the problem. Further, the relative improvement of AFLAC to AFLAC-Abl is more significant for the classes $0 \sim 4$ than for $5 \sim 9$ in BMNISTR-1 and BMNISTR-3, suggesting that AFLAC tends to increase performance more significantly for classes in which the domain-class dependency occurs. Moreover, the improvement is more significant in BMNISTR-1 than in BMNISTR-2, suggesting that the stronger the domain-class dependency is, the lower the performance of domain-invariance-based methods becomes. This result is consistent with Theorem 2 which shows that the trade-off is likely to occur when $I(d; y)$ is large. Finally, although the dependencies of BMNISTR-1 and BMNISTR-3 have different trends, AFLAC improved the F-measures in both datasets.

我们首先调查了领域类依赖对 IFL 方法性能的影响程度。在表 1 右侧，我们比较了 BMNISTR 中类 0 到 4 和类 5 到 9 与目标领域 M0 的平均 F 值。请注意，类 $0 \sim 4$ 的样本大小在不同领域之间是可变的，而类 $5 \sim 9$ 在不同领域之间的样本大小是相同的 (表 1 左侧)。F 值显示，AFLAC 在大多数数据集-类对中优于基线，这支持了领域类依赖降低基于领域不变性的方法性能的观点，并且 AFLAC 可以缓解这个问题。此外，在 BMNISTR-1 和 BMNISTR-3 中，AFLAC 与 AFLAC-Abl 的相对改进在类 $0 \sim 4$ 中比在类 $5 \sim 9$ 中更为显著，这表明 AFLAC 在领域类依赖发生的类中更倾向于显著提高性能。此外，在 BMNISTR-1 中的改进比在 BMNISTR-2 中更为显著，这表明领域类依赖越强，基于领域不变性的方法性能越低。这个结果与定理 2 一致，该定理表明当 $I(d; y)$ 较大时，可能会出现权衡。最后，尽管 BMNISTR-1 和 BMNISTR-3 的依赖趋势不同，AFLAC 在这两个数据集中的 F 值都有所提高。

Table 2. Accuracies for each dataset and target domain. The $I(d; y)$ column is estimated from source domain data, which indicates the domain-class dependency.

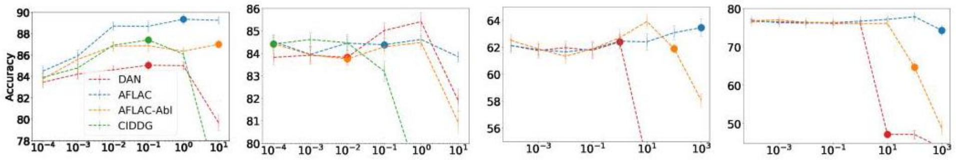
表 2. 每个数据集和目标领域的准确率。$I(d; y)$ 列是从源领域数据中估计的，表示领域类依赖。

Dataset	Target	I(d; y)	CNN	DAN	CIDDG	AFLAC-Abl	AFLAC
BMNISTR-1	M0	0.026	83.9 ± 0.4	85.0 ± 0.4	87.4 ± 0.3	87.0 ± 0.4	89.3 ± 0.4
	M15	0.034	98.5 ± 0.2	98.5 ± 0.1	98.3 ± 0.2	98.3 ± 0.2	98.8 ± 0.1
	M30	0.037	97.5 ± 0.1	97.4 ± 0.1	97.4 ± 0.2	97.6 ± 0.1	98.3 ± 0.2
	M45	0.036	89.9 ± 0.9	90.2 ± 0.6	89.8 ± 0.5	92.8 ± 0.5	93.3 ± 0.6
	M60	0.030	96.7 ± 0.3	97.0 ± 0.2	97.2 ± 0.1	96.6 ± 0.2	97.4 ± 0.2
	M75	0.017	87.1 ± 0.5	87.3 ± 0.4		87.7 ± 0.5	88.1 ± 0.4
	Avg		92.3	92.6	93.1	93.3	94.2
BMNISTR-2 Avg			92.2	92.7	93.1	94.0	94.5
BMNISTR-3 Av			90.6	91.7	91.4	91.6	92.9
PACS	photo	0.102	82.2 ± 0.4	81.8 ± 0.4	-	82.5 ± 0.4	83.5 ± 0.3
	art_painting	0.117	61.0 ± 0.5	60.9 ± 0.5	-	62.6 ± 0.4	63.3 ± 0.3
	cartoon	0.131	64.9 ± 0.5	64.9 ± 0.6	-	64.2 ± 0.3	64.9 ± 0.3
	sketch	0.023	61.4 ± 0.5	61.4 ± 0.5	-	59.6 ± 0.7	60.1 ± 0.7
	Avg		67.4	67.2	-	67.2	68.0
WISDM	16 users	0.181	84.0 ± 0.4	83.8 ± 0.3	84.4 ± 0.483.7 ± 0.3		84.4 ± 0.3
IEMOCAP	Ses01F	0.005	56.0 ± 0.7	60.1 ± 0.7	-	62.9 ± 0.5	60.4 ± 0.9
	Ses01M		61.0 ± 0.3	63.5 ± 0.5	-	68.0 ± 0.5	66.1 ± 0.3
	Ses02F	0.045	61.2 ± 0.5	60.4 ± 0.5	-	65.8 ± 0.5	64.2 ± 0.4
	Ses02M		76.6 ± 0.4	47.2 ± 0.7	-	64.7 ± 1.7	74.3 ± 1.3
	Ses03F	0.037	69.2 ± 0.9	71.9 ± 0.4	-	70.0 ± 0.6	70.1 ± 0.4
	Ses03M		56.9 ± 0.4	57.3 ± 0.5	-	56.2 ± 0.4	56.8 ± 0.4
	Ses04F	0.120	75.5 ± 0.5	75.5 ± 0.6	-	75.4 ± 0.6	75.7 ±0.6
	Ses04M		58.5 ± 0.5	57.4 ± 0.5	-	58.7 ± 0.5	59.2 ± 0.5
	Ses05F	0.063	61.8 ± 0.4	62.4 ± 0.5	-	61.9 ± 0.3	63.4 ± 0.7
	Ses05M		47.6 ± 0.3	46.9 ± 0.4	-	49.6 ± 0.4	49.9 ± 0.4
	Avg		62.4	60.3	-	63.3	64.0

数据集	目标	I(d; y)	CNN	DAN	CIDDG	AFLAC-Abl	AFLAC
BMNISTR-1	M0	0.026	83.9 ± 0.4	85.0 ± 0.4	87.4 ± 0.3	87.0 ± 0.4	89.3 ± 0.4
	M15	0.034	98.5 ± 0.2	98.5 ± 0.1	98.3 ± 0.2	98.3 ± 0.2	98.8 ± 0.1
	M30	0.037	97.5 ± 0.1	97.4 ± 0.1	97.4 ± 0.2	97.6 ± 0.1	98.3 ± 0.2
	M45	0.036	89.9 ± 0.9	90.2 ± 0.6	89.8 ± 0.5	92.8 ± 0.5	93.3 ± 0.6
	M60	0.030	96.7 ± 0.3	97.0 ± 0.2	97.2 ± 0.1	96.6 ± 0.2	97.4 ± 0.2
	M75	0.017	87.1 ± 0.5	87.3 ± 0.4		87.7 ± 0.5	88.1 ± 0.4
	平均		92.3	92.6	93.1	93.3	94.2
BMNISTR-2 平均			92.2	92.7	93.1	94.0	94.5
BMNISTR-3 平均			90.6	91.7	91.4	91.6	92.9
PACS	照片	0.102	82.2 ± 0.4	81.8 ± 0.4	-	82.5 ± 0.4	83.5 ± 0.3
	艺术 _ 绘画	0.117	61.0 ± 0.5	60.9 ± 0.5	-	62.6 ± 0.4	63.3 ± 0.3
	卡通	0.131	64.9 ± 0.5	64.9 ± 0.6	-	64.2 ± 0.3	64.9 ± 0.3
	草图	0.023	61.4 ± 0.5	61.4 ± 0.5	-	59.6 ± 0.7	60.1 ± 0.7
	平均		67.4	67.2	-	67.2	68.0
WISDM	16 名用户	0.181	84.0 ± 0.4	83.8 ± 0.3	84.4 ± 0.483.7 ± 0.3		84.4 ± 0.3
IEMOCAP	Ses01F	0.005	56.0 ± 0.7	60.1 ± 0.7	-	62.9 ± 0.5	60.4 ± 0.9
	Ses01M		61.0 ± 0.3	63.5 ± 0.5	-	68.0 ± 0.5	66.1 ± 0.3
	Ses02F	0.045	61.2 ± 0.5	60.4 ± 0.5	-	65.8 ± 0.5	64.2 ± 0.4
	Ses02M		76.6 ± 0.4	47.2 ± 0.7	-	64.7 ± 1.7	74.3 ± 1.3
	Ses03F	0.037	69.2 ± 0.9	71.9 ± 0.4	-	70.0 ± 0.6	70.1 ± 0.4
	Ses03M		56.9 ± 0.4	57.3 ± 0.5	-	56.2 ± 0.4	56.8 ± 0.4
	Ses04F	0.120	75.5 ± 0.5	75.5 ± 0.6	-	75.4 ± 0.6	75.7 ±0.6
	Ses04M		58.5 ± 0.5	57.4 ± 0.5	-	58.7 ± 0.5	59.2 ± 0.5
	Ses05F	0.063	61.8 ± 0.4	62.4 ± 0.5	-	61.9 ± 0.3	63.4 ± 0.7
	Ses05M		47.6 ± 0.3	46.9 ± 0.4	-	49.6 ± 0.4	49.9 ± 0.4
	平均		62.4	60.3	-	63.3	64.0

Next we compared the mean accuracies (with standard errors) in both synthetic (BMNISTR) and real-world (PACS, WISDM, and IEMOCAP) datasets (Table 2). Note that the performance of our baseline CNN on PACS, WISDM, and IEMOCAP is similar but partly different from that reported in previous studies ([22], [1], and [8], respectively) probably because the DG performance strongly depends on validation methods and other implementation details as reported in many recent studies [18 2 23]. Also, we trained CIDDG only on BMNISTR and WISDM due to computational resource constraint. This table enables us to make the following observations. (1) Domain-class dependency in real-world datasets negatively affects the DG performance of IFL methods. The results obtained on PACS (Avg) and WISDM showed that the vanilla CNN outperformed the IFL methods (DAN and AFLAC-Abl). Additionally, the results on IEMOCAP shows that AFLAC tended to outperform AFLAC-Abl when $I\left(d;y\right)$ had large values (in Ses04 and Ses05), which is again consistent with Theorem 2 . These results support the importance of considering domain-class dependency in real-world datasets. (2) AFLAC performed better than the baselines on all the datasets in average, except for CIDDG on WISDM. Note that AFLAC is more parameter efficient than CIDDG as we noted in Sec. 2.2. These results supports the efficacy of the proposed model to overcome the trade-off problem.

接下来，我们比较了合成数据集 (BMNISTR) 和真实世界数据集 (PACS、WISDM 和 IEMOCAP) 中的平均准确率 (带标准误差)(表 2)。注意，我们的基线 CNN 在 PACS、WISDM 和 IEMOCAP 上的表现相似，但与之前研究中报告的结果 (分别为 [22]、[1] 和 [8]) 部分不同，这可能是因为 DG 性能在很大程度上依赖于验证方法和其他实现细节，正如许多近期研究所报告的那样 [18 2 23]。此外，由于计算资源的限制，我们仅在 BMNISTR 和 WISDM 上训练了 CIDDG。该表使我们能够得出以下观察结果。(1) 真实世界数据集中的领域类别依赖性对 IFL 方法的 DG 性能产生了负面影响。在 PACS(平均值) 和 WISDM 上获得的结果表明，普通 CNN 的表现优于 IFL 方法 (DAN 和 AFLAC-Abl)。此外，IEMOCAP 上的结果显示，当 $I(d;y)$ 值较大时 (在 Ses04 和 Ses05 中)，AFLAC 的表现往往优于 AFLAC-Abl，这与定理 2 一致。这些结果支持在真实世界数据集中考虑领域类别依赖性的重要性。(2)AFLAC 在所有数据集上的平均表现优于基线，除了 WISDM 上的 CIDDG。注意，正如我们在第 2.2 节中提到的，AFLAC 比 CIDDG 更加参数高效。这些结果支持所提模型克服权衡问题的有效性。



(a) BMNISTR-1, M0 (b) WISDM (c) IEMOCAP, 02M (d) IEMOCAP, 05F

Fig. 3. Classification Accuracy with various γ . Each caption shows dataset name and target domain. The round markers correspond to γ values chosen by validation. The error bars correspond to standard errors.

图 3. 不同 γ 的分类准确率。每个标题显示数据集名称和目标领域。圆形标记对应于通过验证选择的 γ 值。误差条对应于标准误差。

Finally, we investigated the relationship between the strength of regularization and performance. In DG, it is difficult to choose appropriate hyperparameters because we cannot use target domain data at valiadtion step (since they are not available during training); therefore, hyperparameter insensitivity is significant in DG. Figure 3 shows the hyperparameter sensitivity of the classification accuracies for DAN, CIDDG, AFLAC-Abl, and AFLAC. These figures suggest that DAN and AFLAC-Abl sometimes outperformed AFLAC with appropriate γ values, but there is no guarantee that such γ values will be chosen by validation whereas AFLAC is robust toward hyperparameter choice. Specifically, as shown in Figures $3-(\mathrm{b}, \mathrm{d})$, DAN and AFLAC-Abl outperformed AFLAC with $\gamma=1$ and 10, respectively. One possible explanation of those results is that accuracy for target domain sometimes improves by giving priority to domain invariance at the cost of the accuracies for source domains, but AFLAC improves domain invariance only within a range that does not interfere with accuracy for source domains. However, as shown in Figure 3, the performance of DAN and AFLAC-Abl are sensitive to hyperparameter choice. For example, although they got high scores with $\gamma=1$ in Figure 3-(b), the scores dropped rappidly when γ increases to 10 or decreases to 0.01 . Also, the scores of DAN and AFLAC-Abl in Figure 3-(c) dropped significantly with $\gamma>10$, and such large γ was indeed chosen by overfitting to validation domain. On the other hand, Figures 3-(a, b, c, d) show that the accuracy gaps of AFLAC-Abl and AFLAC increase with strong regularization (such as when $\gamma=10$ or 100). These results suggest that AFLAC, as it was designed, does not tend to reduce the classification accuracy with strong regularizer, and such robustness of AFLAC might have yileded the best performance shown in Table 2.

最后，我们研究了正则化强度与性能之间的关系。在领域泛化 (DG) 中，由于在验证步骤中无法使用目标领域数据 (因为在训练期间不可用)，因此选择适当的超参数是困难的；因此，超参数的不敏感性在 DG 中显著。图 3 显示了 DAN、CIDDG、AFLAC-Abl 和 AFLAC 的分类准确率的超参数敏感性。这些图表表明，DAN 和 AFLAC-Abl 在适当的 γ 值下有时优于 AFLAC，但并不能保证这些 γ 值会通过验证被选择，而 AFLAC 对超参数选择具有鲁棒性。具体而言，如图 $3-(\mathrm{b}, \mathrm{d})$ 所示，DAN 和 AFLAC-Abl 在 $\gamma=1$ 和 10 时分别优于 AFLAC。这些结果的一个可能解释是，目标领域的准确率有时通过优先考虑领域不变性而改善，代价是源领域的准确率，但 AFLAC 仅在不干扰源领域准确率的范围内改善领域不变性。然而，如图 3 所示，DAN 和 AFLAC-Abl 的性能对超参数选择敏感。例如，尽管它们在图 3-(b) 中以 $\gamma=1$ 获得了高分，但当 γ 增加到 10 或减少到 0.01 时，分数迅速下降。此外，图 3-(c) 中 DAN 和 AFLAC-Abl 的分数在 $\gamma>10$ 下显著下降，这样大的 γ 确实是通过对验证领域的过拟合而选择的。另一方面，图 3-(a, b, c, d) 显示，AFLAC-Abl 和 AFLAC 的准确率差距随着强正则化 (例如，当 $\gamma=10$ 或

100 时) 而增加。这些结果表明，AFLAC 在设计时并不倾向于在强正则化下降低分类准确率，而 AFLAC 的这种鲁棒性可能导致了表 2 中所示的最佳性能。

5 Conclusion

5 结论

In this paper, we addressed domain generalization under domain-class dependency, which was overlooked by most prior DG methods relying on IFL. We theoretically showed the importance of considering the dependency and the way to overcome the problem by expanding the analysis of [33]. We then proposed a novel method AFLAC, which maximizes domain invariance within a range that does not interfere with classification accuracy on adversarial training. Empirical validations show the superior performance of AFLAC to the baseline methods, supporting the importance of the domain-class dependency in DG tasks and the efficacy of the proposed method to overcome the issue.

在本文中，我们探讨了在领域类依赖下的领域泛化问题，这一问题在大多数依赖于 IFL 的先前 DG 方法中被忽视。我们从理论上展示了考虑这种依赖的重要性，以及通过扩展 [33] 的分析来克服该问题的方法。随后，我们提出了一种新方法 AFLAC，该方法在不干扰对抗训练分类准确性的范围内最大化领域不变性。实证验证表明，AFLAC 的性能优于基线方法，支持了领域类依赖在 DG 任务中的重要性以及所提方法克服该问题的有效性。

References

参考文献

1. Andrey, I.: Real-time human activity recognition from accelerometer data using convolutional neural networks. Applied Soft Computing (2017)

2. Balaji, Y., Sankaranarayanan, S., Chellappa, R.: Metareg: Towards domain generalization using meta-regularization. In: Advances in Neural Information Processing Systems 31 (2018)

3. Blanchard, G., Lee, G., Scott, C.: Generalizing from several related classification tasks to a new unlabeled sample. In: Proc. of the 24th International Conference on Neural Information Processing Systems (2011)

4. Busso, C., Bulut, M., Lee, C.C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J.N., Lee, S., Narayanan, S.S.: Iemocap: interactive emotional dyadic motion capture database. Language Resources and Evaluation **42** (4) , 335 (Nov 2008)

5. Chen, M., He, X., Yang, J., Zhang, H.: 3-d convolutional recurrent neural networks with attention model for speech emotion recognition. IEEE Signal Processing Letters 25, 1-1 (07 2018)

6. Chou, J.C., chieh Yeh, C., yi Lee, H., shan Lee, L.: Multi-target voice conversion without parallel data by adversarially learning disentangled audio representations. In: Proc. Interspeech (2018)

7. Erfani, S., Baktashmotlagh, M., Moshtaghi, M., Nguyen, V., Leckie, C., Bailey, J., Kotagiri, R.: Robust domain generalisation by enforcing distribution invariance. In: 25th International Joint Conference on Artificial Intelligence (2016)

8. Etienne, C., Fidanza, G., Petrovskii, A., Devillers, L., Schmauch, B.: Speech emotion recognition with data augmentation and layer-wise learning rate adjustment. CoRR abs/1802.05630 (2018), http://arxiv.org/abs/1802.05630

9. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. J. Mach. Learn. Res. (2016)

10. Ghifary, M., Balduzzi, D., Kleijn, W.B., Zhang, M.: Scatter component analysis: A unified framework for domain adaptation and domain generalization. IEEE Transactions on Pattern Analysis and Machine Intelligence (2017)

11. Ghifary, M., Bastiaan Kleijn, W., Zhang, M., Balduzzi, D.: Domain generalization for object recognition with multi-task autoencoders. In: Proc. of the IEEE International Conference on Computer Vision (ICCV) (2015)

12. Gong, M., Zhang, K., Liu, T., Tao, D., Glymour, C., Schölkopf, B.: Domain adaptation with conditional transferable components. In: Proc. of the 33rd International Conference on International Conference on Machine Learning (2016)

13. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Proc. of the 27th International Conference on Neural Information Processing Systems (2014)

14. Iwasawa, Y., Nakayama, K., Yairi, I., Matsuo, Y.: Privacy issues regarding the application of dnns to activity-recognition using wearables and its countermeasures by use of adversarial training. In: Proc. of the 26th International Joint Conference on Artificial Intelligence. pp. 1930-1936 (2017)

15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Proc. of the 25th International Conference on Neural Information Processing Systems. pp. 1097-1105 (2012)

16. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Activity recognition using cell phone accelerometers. SIGKDD Explor. Newsl. (2011)

17. Lample, G., Zeghidour, N., Usunier, N., Bordes, A., Denoyer, L., Ranzato, M.: Fader networks:manipulating images by sliding attributes. In: Proc. of the 30th Neural Information Processing Systems (2017)

18. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Deeper, broader and artier domain generalization. In: Proc. of the IEEE International Conference on Computer Vision (ICCV) (2017)

19. Li, D., Yang, Y., Song, Y., Hospedales, T.M.: Learning to generalize: Meta-learning for domain generalization. In: Proc. of the 32nd AAAI Conference on Artificial Intelligence (2018)

20. Li, H., Jialin Pan, S., Wang, S., Kot, A.C.: Domain generalization with adversarial feature learning. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)

21. Li, Y., Gong, M., Tian, X., Liu, T., Tao, D.: Domain generalization via conditional invariant representations. In: Proc. of the 32nd AAAI Conference on Artificial Intelligence (2018)

22. Li, Y., Tian, X., Gong, M., Liu, Y., Liu, T., Zhang, K., Tao, D.: Deep domain generalization via conditional invariant adversarial networks. In: The European Conference on Computer Vision (ECCV) (September 2018)

23. Li, Y., Yang, Y., Zhou, W., Hospedales, T.M.: Feature-critic networks for heterogeneous domain generalization. CoRR abs/1901.11448 (2019), http://arxiv.org/ abs/1901.11448

24. Louizos, C., Swersky, K., Li, Y., Welling, M., Zemel, R.S.: The variational fair autoencoder. In: Proc. International Conference on Representation Learning (2016)

25. Madras, D., Creager, E., Pitassi, T., Zemel, R.S.: Learning adversarially fair and transferable representations. In: Proc. of the 35th International Conference on Machine Learning (2018)

26. Motiian, S., Piccirilli, M., Adjeroh, D.A., Doretto, G.: Unified deep supervised domain adaptation and generalization. In: Proc. of the IEEE International Conference on Computer Vision (ICCV) (2017)

27. Muandet, K., Balduzzi, D., Schlkopf, B.: Domain generalization via invariant feature representation. In: Proc. of the 30th International Conference on Machine Learning (2013)

28. Shankar, S., Piratla, V., Chakrabarti, S., Chaudhuri, S., Jyothi, P., Sarawagi, S.: Generalizing across domains via cross-gradient training. In: Proc. International Conference on Learning Representations (2018)

29. Sriram, A., Jun, H., Gaur, Y., Satheesh, S.: Robust speech recognition using generative adversarial networks. In: The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2018)

30. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (2011)

31. Tzeng, E., Hoffman, J., Darrell, T., Saenko, K.: Simultaneous deep transfer across domains and tasks. In: Proc. of the IEEE International Conference on Computer Vision (ICCV) (2015)

32. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.: Deep domain confusion: Maximizing for domain invariance. CoRR abs/1412.3474 (2014), http://arxiv org/abs/1412.3474

33. Xie, Q., Dai, Z., Du, Y., Hovy, E., Neubig, G.: Controllable invariance through adversarial feature learning. In: Proc. of the 30th International Conference on Neural Information Processing Systems (2017)

34. Yang, J., Nguyen, M.N., San, P.P., Li, X., Krishnaswamy, S.: Deep convolutional neural networks on multichannel time series for human activity recognition. In: Proc. of the 24th International Joint Conference on Artificial Intelligence (2015)

35. Zemel, R., Wu, Y., Swersky, K., Pitassi, T., Dwork, C.: Learning fair representations. In: Proc. of the 30th International Conference on Machine Learning (2013)

36. Zhang, K., Schlkopf, B., Muandet, K., Wang, Z.: Domain adaptation under target and conditional shift. In: Proc. of the 30th International Conference on Machine Learning (2013)

37. M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi.: Learning sleep stages from radio signals: A conditional adversarial architecture. In: Proc. of the 34th International Conference on