数据生成也是一种流行的技术，用于生成多样化的训练数据，以提高模型的泛化能力。数据生成通常使用一些生成模型来实现，如变分自动编码器（VAE）[14]和生成对抗性网络（GAN）[15]。此外，它还可以使用Mixup[16]策略来实现。Rahman等人[56]使用ComboGAN[175]生成新数据，然后应用领域差异度量（如MMD[176]）来最小化真实图像和生成图像之间的分布差异，以帮助学习一般表示。Qiao等人[40]利用对抗性训练来创建“虚构”但“具有挑战性”的群体，其中Wasserstein自动编码器（WAE）[177]用于帮助生成保留语义并具有大域传输的样本。周等人[43]在语义一致的情况下生成新分布，然后最大化源分布和新分布之间的差异。Somavarapu等人[57]引入了一种基于图像风格化的简单变换，以探索跨源可变性，从而获得更好的泛化效果，其中AdaIN[178]用于实现对任意风格的快速风格化。与其他人不同的是，李等人[52]使用对抗性训练来生成域，而不是样本。由于涉及不同的生成模型，这些方法更为复杂，我们应该注意模型容量和计算成本。除了上述生成模型外，Mixup[174]也是一种流行的数据生成技术。Mixup通过在任意两个实例之间以及它们的标签之间执行线性插值来生成新数据，该插值具有从贝塔分布采样的权重，这不需要训练生成模型。最近，有几种方法对DG使用Mixup，通过在原始空间[47，48，54]中执行Mixup来生成新的样本；或者在不显式地生成原始训练样本的特征空间[491948，55]中。这些方法在保持概念和计算简单的同时，在流行的基准测试上实现了有希望的性能。