SIDA: 合成图像驱动的零样本领域适应

Ye-Chan Kim
dpcksd178@hanyang.ac.kr
Hanyang University

SeungJu Cha
sju9020@hanyang.ac.kr
Hanyang University

Si-Woo Kim
boreng0817@hanyang.ac.kr
Hanyang University

Taewhan Kim
taewhan@hanyang.ac.kr
Hanyang University

Dong-Jin Kim[†]
djdkim@hanyang.ac.kr
Hanyang University

Abstract

零样本领域适应是一种在不使用目标域图像数据的情况下使模型适应目标域的方法。为了在没有目标图像的情况下进行适应，现有研究利用 CLIP 的嵌入空间和文本描述来模拟类似目标的风格特征。尽管之前为零样本领域适应中取得了一些成就，但我们发现这些文本驱动的方法在捕捉复杂的真实世界变化时表现不佳，并且由于它们的对齐过程，显著增加了适应时间。我们探索利用图像数据的解决方案，而不是依赖文本描述，图像数据提供了多样且更加细粒度的风格提示。在这项工作中，我们提出了 SIDA，这是一种新颖且高效的利用合成图像的零样本领域适应方法。为了生成合成图像，我们首先创建详细的、类似源域的图像，并应用图像翻译以反映目标域的风格。然后，我们利用这些合成图像的风格特征作为目标域的代理。基于这些特征，我们引入了域混合和贴片风格转移模块，使得真实世界变化的有效建模成为可能。特别地，域混合通过混合多种风格来扩展域内表示，而贴片风格转移则为各个贴片分配不同的风格。我们通过在多样的零样本适应场景中，特别是在具有挑战性的域中，展示领先水平的表现，以证明我们方法的有效性。此外，我们的方法通过显著减少总体适应时间，实现了高效能。

CCS Concepts

• Computing methodologies → Scene understanding; Transfer learning.

Keywords

Zero-Shot Domain Adaptation, Synthetic Data, Feature Style Transfer

为了应对深度学习中固有的领域迁移问题，基于无监督领域自适应 (UDA) 的各种领域自适应 (DA) 研究被提出。然而，由于目标领域图像数据数量有限，这些方法常常难以在危险或罕见场景 (例如，“火灾”和“沙尘暴”) 中进行适应。为了解决这一限制，不依赖任何目标领域图像数据的零样本领域自适应 (ZSDA) 出现了。然而，缺乏目标图像在有效的领域适应中引入了相当大的挑战。因此，大多数现有的方法依赖于文本描述来近似目标领域特征。例如，P DA 和 ULDA 利用 CLIP 的语言嵌入能力来优化目标领域风格的可学习向量。然后他们在特征空间中应用风格转移算法，使用优化后的向量以获得类似目标领域特征，而不需要目标领域的图像数据。

然而，尽管 P $\emptyset$ DA [?] 和 ULDA [?] 有所改进，在利用文本描述来制作目标风格特征时仍有若干关键局限。首先，简单固定的文本描述 (例如，“在 {域} 驾驶。”) 来模拟目标风格不足以反映真实世界的多样环境。在图 ?? 中，我们展示了同一

领域 (a) 和 (b) 内的整体风格强度变化以及单一图像中的局部风格变化。尽管有这些变化，现有方法忽视了在现实场景中常见的风格强度差异。它们通常依赖于固定文本描述和单一模拟的目标风格特征。其次，文本驱动的方法需要对每一个源图像进行额外的对齐过程，这个过程旨在优化应用于图像特征的可学习向量。随着源数据规模的增长，这个对齐过程变得低效。

为了克服这些限制，我们的目标是在图像领域推导出类似目标的风格特征。我们注意到最近生成模型 [????] 有效生成多样且与真实世界相似的图像的能力。在此基础上，我们提出了 SIDA (合成图像驱动的零样本领域适应)，这是一种新颖且高效的零样本领域适应方法，可有效利用来自合成图像的类似目标的风格特征。SIDA 的架构包含三个关键阶段：图像生成过程、域混合 & 贴片风格转移模块和微调阶段。首先，我们通过基于 VLM 的字幕生成器 [?] 从源图像中提取场景描述。根据提取的描述，我们通过图像生成器 [?] 生成各种类似源的合成图像，反映真实源图像的内容。之后，我们使用图像翻译模型 [?] 将类似源的合成图像翻译成各种目标域。

然后，我们引入 (1) 域混合和 (2) 块风格迁移来丰富全局和局部风格的变化。域混合旨在通过模拟全局风格强度的变化来获取多样的目标风格。具体而言，我们将来自主要目标域和辅助域的合成图像的风格特征进行混合。此外，为了帮助模型捕捉局部风格强度的变化，我们引入块风格迁移，将不同的目标风格特征应用于每个块。我们利用域混合产生的多样目标风格，在每个块的基础上具有变化的强度。最后，我们通过引入基于熵信息的加权交叉熵损失函数来微调模型，以有效学习新风格的适当权重。

我们在各种现有的 ZSDA 实验环境中验证了 SIDA。我们的方法在多个场景中始终表现出优越的性能和效率。特别是，SIDA 在火灾和沙尘暴等具有挑战性的环境中显示出改进的性能，这些环境中由于数据的稀缺性 ZSDA 尤为重要。

总之，我们的贡献如下：

- 我们提出了 SIDA，这是一种高效且有效的 ZSDA 方法，它利用来自图像域的合成图像，而不是依赖于文本描述。
- 我们提出了域混合和局部风格转移方法，可以模拟出大量多样化的全局风格强度和局部风格变化，如同真实世界情境。
- 大量实验表明，结合额外熵信息的 SIDA 在各种 ZSDA 设置中实现了显著的性能提升。

1 相关工作

零样本领域适应。与依赖目标领域图像数据的对抗领域适应 (UDA) 和领域泛化 (DG)[??] 不同，最近的 ZSDA 方法 [??] 利用 CLIP 嵌入 [?] 通过文本驱动的风格转换生成目标适应特

[†]Corresponding author.