MeGA-CDA: Memory Guided Attention for Category-Aware Unsupervised Domain Adaptive Object Detection

MeGA-CDA：记忆引导注意的类别感知无监督领域自适应目标检测

Abstract

现有的无监督域自适应目标检测方法通过对抗性训练进行特征对齐。虽然这些方法在性能上实现了合理的改进，但它们通常执行类别无关的领域对齐，从而导致特征的负迁移。为了解决这一问题，在本研究中，我们试图通过提出分类感知域适应(Memory Guided Attention for Category-Aware Domain Adaptation：MeGA-CDA)的记忆指导注意，将类别信息整合到域适应过程中。该方法采用类别识别器来保证类别感知特征对齐，以学习领域不变的识别特征。然而，由于目标样本无法获得类别信息，我们建议生成记忆引导的类别特定注意图，然后用于将特征适当地路由到相应的类别识别器（discriminator）。在多个基准数据集上对该方法进行了评估，结果表明该方法优于现有方法。

1. Introduction

现有方法的问题：泛化性能一般，与训练图像相比，这些方法在从不同分布采样的图像上进行评估时，会遭受严重的性能退化。因此，开发能够更好地推广检测器的方法是至关重要的。

解决这一问题的一种方法是通过**无监督域自适应：**利用带标记的源域数据和未带标记的目标域数据来适应目标检测器，并提高对无标记目标域数据的性能。为了解决这个问题，通常这些方法试图通过在源图像和目标图像之间进行特征对齐来学习域不变特征。基于最小化域间差异以减小目标域的上限误差的理论思想，通过对抗性训练实现特征匹配。虽然这些方法带来了相当大的改进，但它们以不可知的分类方式执行域对齐。也就是说，它们在不考虑类别信息的情况下匹配两个域的全局边际分布。这可能导致目标域样本与不同类别的源域样本不正确对齐的情况(见图 1) ，从而导致次优的适应性能。由于存在多种类型的物体，因此适应物体检测器的任务特别容易出现这种问题。

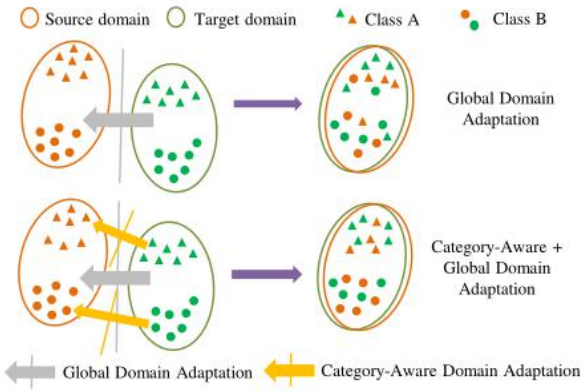

Figure 1. Performing global domain adaptation alone results in potential negative transfer of features. To mitigate this issue, we employ additional category-aware adaptation.

考虑到这个问题，我们通过匹配特征的局部联合分布以及全局比对，将类别信息纳入域自适应过程。特别是，我们通过在训练过程中使用特定类别的鉴别器来对特征进行分类比对。请注意，这需要逐像素的类别标签，以便特征可以显式路由到各自的类别特定的鉴别器。然而，在非监督域自适应的情况下，注释不能以任何形式(边界框/类别标签/逐像素标签)为目标数据集提供。由于缺乏注释，因此很难使用特定类别的识别器。

为了克服这一挑战，我们提出记忆引导的注意图来实现对类别感知（category-aware）特征的对齐。这些注意力图的目标是聚焦于特定类别的对象，因此可以用来将主干特征路由到适当的类别特定识别器。为了生成这些注意力地图，我们建议使用记忆网络。记忆网络的使用灵感来自于它们在更长时间内存储模式的能力。此外，使用显式写操作更新模式的能力使它们在领域对抗训练中特别有用，因为特性会随着训练过程的变化而变化。为了确定特定位置的注意力，我们使用该位置的特征作为查询，从不同类别特定的记忆网络中检索相关项目。然后将检索到的条目与查询条目进行比较，并基于相似度计算特定类别的注意图。此外，为了提高记忆模块和注意图生成过程的有效性，我们提出了一种基于度量学习的方法，该方法包括基于源域中可用的弱监督来学习适当的相似性度量。为了证明该方法的有效性，我们在几个基准数据集和自适应协议上对其进行了评估。此外，我们还发现记忆引导的注意图在实现类别分布匹配方面发挥了重要作用，从而减少了错误的特征对齐问题。综上所述，我们工作的主要贡献如下:

我们提出记忆引导的注意图，使分类分布匹配领域自适应目标检测。

此外，我们利用基于度量学习的方法来计算特定类别的注意图，从而提高记忆模块的有效性。

该方法在几个基准数据集上进行了评估，并显示出比最近的领域自适应检测方法有相当大的优势。此外，我们进行了详细的消融研究，以明确消除记忆引导注意在实现类别对齐中的作用。

3.Method

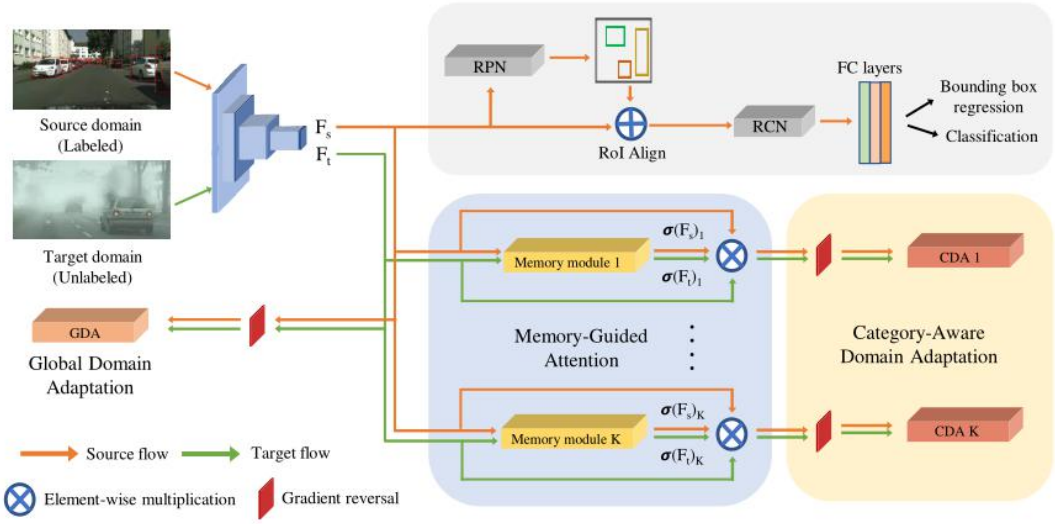

图 2。方法概述。通过全局域适应和类别感知域适应实现源特征和目标特征的对齐。全局对齐是通过类别不可知的全局标识符实现的，而类别感知的对齐是通过使用 K 个类别特定的标识符实现的。由于目标标签不可用，这些鉴别器的特征使用记忆引导的类别特定注意图进行路由。

我们假设可用的完全标记的源域图像带有边界框注释和未标记的目标域图像没有任何注释。

源数据集：$D_s = \{X_s^i, b_s^i, y_s^i\}_{i=1}^{N_s}$，$X_s^i$：第 i 张图像，$b_s^i$ y_s^i 表示包围框注释，并对应在第 i 张源域图像中的类别标签。此外，每个类别标签表明一个目标存在于数据集和一个额外的类别为背景类，此外，我们将目标域数据集表示为$D_t = \{X_t^i\}_{i=1}^{N_t}$，其中 X_t^i 表示第 i 个目标域图像。我们使用 fast - rcnn 作为我们的基本模型。我们表示检测网络的骨干特征编码器为ε。该方法的目标是利用源域标签信息学习一个对目标域图像具有良好性能的检测网络。为了实现这一目标，我们采用特征对齐方法，通过域对抗训练，对源数据集和目标数据集的图像进行特征编码网络提取的特征分布进行匹配。

Fig.2 提出功能对齐方法的概述包括三个主要模块:1) 全局鉴别器，对由特征编码器网络提取的整个特征映射进行校准 2) 分类明智的识别器，专注于各自的类别特定的信息，以在源和目标域之间对齐属于对应类别的特征。3)记忆引导注意机制，通过在提取的特征地图上生成类别特定注意，实现分类智能判别器的训练。这种注意有助于集中在提取的特征图中的类别信息，以训练各自的类别甄别器。注意力是通过特定类别的记忆模块产生的，该模块存储了相应对象类别的相关信息。这些模块的特性对齐细节将在下面几节中描述。

3.1 全局自适应鉴别器

在现有工作的基础上，我们还采用了一个全局鉴别器，在图像级别对特征映射进行特征

对齐。全局鉴别器($\mathcal{D}_{gda}$)接受从骨干网络中提取的全部特征映射，并训练识别特征映射是从源域还是目标域提取。更准确地说，让我们表示一个特征图 $F_s,\ F_t\ \in \mathbb{R}^{C\times H\times W}$ 分别从任何源文件 X_s 和目标域文件 X_t 中提取。利用全局判别函数 $\mathcal{D}_{gda}$ 提供了一个 H×W 的预测图，并利用最小二乘损失监督训练判别器网络，其中域标签 $y_d\ \in\ 0,1.$，对于源数据, $\forall X_s \in D_s$ 和目标数据, $\forall X_t \in D_t$，域标签分别设置为 1 和 0。整个损失函数可以写成:

$$\mathcal{L}_{gda}(X_s, X_t) = -\sum_{h=1}^{H}\sum_{w=1}^{W} y_d(1-\mathcal{D}_{gda}(F_s^{(h,w)}))^2 + (1-y_d)(\mathcal{D}_{gda}(F_t^{(h,w)}))^2, \quad (1)$$

为了匹配源域特征和目标域特征的分布，我们利用了《Unsupervised domain adaptation by backpropagation》中提出的梯度反转层。梯度反转层在将梯度传播回特征提取网络之前翻转梯度符号。因此，训练鉴别器网络 $\mathcal{D}_{gda}$ 最小化（1），训练特征编码器网络最大化（1）。这种特征提取器和鉴别器之间的对抗训练有助于缩小源图像特征和目标图像特征之间的区域差距。此外，我们利用[32]中提出的最小二乘损失法，而不是利用二元交叉熵损失进行训练，因为它在实践中表现得更好，有助于稳定训练过程。然而，正如我们前面提到的，全局自适应是一种与类别无关的方法，用于在源域和目标域之间执行特征对齐。因此，这导致了特征的负迁移，并损害了整体性能。因此，单独使用全局鉴别器并不是最优的，需要额外的策略来避免特征的负迁移。

3.2. 适应性分类鉴别器

正如前面部分所讨论的，现有的方法只考虑全局特征对齐策略。就目标检测而言，每幅图像可能包含多个类别，因此从这些图像中提取的特征地图将具有属于这些类别的特征，包括背景特征。因此，在源域和目标域对齐时，如何解决类别之间的特征负迁移仍然是域自适应目标检测中的一个重要问题。我们通过使用类别鉴别器(CDA)来解决这个问题，这种鉴别器专注于在源域和目标域之间对齐特定类别的特征。具体来说，我们使用了分类鉴别器，每个鉴别器都专注于对齐各自的类别。

让我们把第 k 类鉴别器表示为 $\mathcal{D}_{cda}^{k}$，F_s、F_t 为分别从源域图像（Xs）和目标域图像（Xt）中提取的特征。为了对齐源域和目标域之间的 第 k 类特征，我们生成了关注映射 $\sigma(F_s)_k,\ \sigma(F_t)_k\ \in \{0,1\}^{H\times W}$ 见第 3.3 节) ，以关注与 第 k 类 相关的信息。类别适应的损失函数可以写成:

$$\mathcal{L}_{cda}^{k}(X_s, X_t) = -\sum_{h=1}^{H}\sum_{w=1}^{W} y_d(1-\mathcal{D}_{cda}^{k}(\sigma(F_s)_k^{(h,w)} \cdot F_s^{(h,w)}))^2 + (1-y_d)(\mathcal{D}_{cda}^{k}(\sigma(F_t)_k^{(h,w)} \cdot F_t^{(h,w)}))^2, \quad (2)$$

其中，$\sigma(\cdot)_k^{(h,w)} = 1$ and $\sigma(\cdot)_k^{(h,w)} = 0$表示 存在和不存在，分别在相应特征图（Fs、Ft）的（h,w）位置处的第 k 类特征。使用这些分类鉴别器进行训练的主要挑战是缺乏关于特征映射中分类位置的信息，特别是对于目标域数据。为此，我们提出了一种通过记忆模块来预测每个类别的注意图位置的机制。

3.3. 记忆引导注意机制

我们提出记忆引导注意机制来帮助类别鉴别器在源域和目标域之间对齐特定类别的特征。具体来说，我们使用与 k 类相对应的 K 个记忆模块。这些存储模块用于在训练过程中存储不同对象的类原型，以便检索这些类原型来计算特定类别的注意图。接下来，我们描述关于内存更新和注意力计算的细节。

3.3.1 Memory module 记忆力模块

记忆力模块有两个操作，即写和读。为了写入存储器，从神经网络中提取的特征被用来适当地更新存储器元件。然而，从神经网络中提取特征，利用存储器读取操作对存储器进行查询，获取最相似的存储元素(或原型特征)。这些操作如图 3 所示。在该方法中，我们学习 K 个记忆模块，即$M_k \in \mathbb{R}^{N_m \times C}$，对应源域和目标域的 K 类，。这里，$N_m$表示每个类别的内存条目数，c 表示特征映射中的通道数。

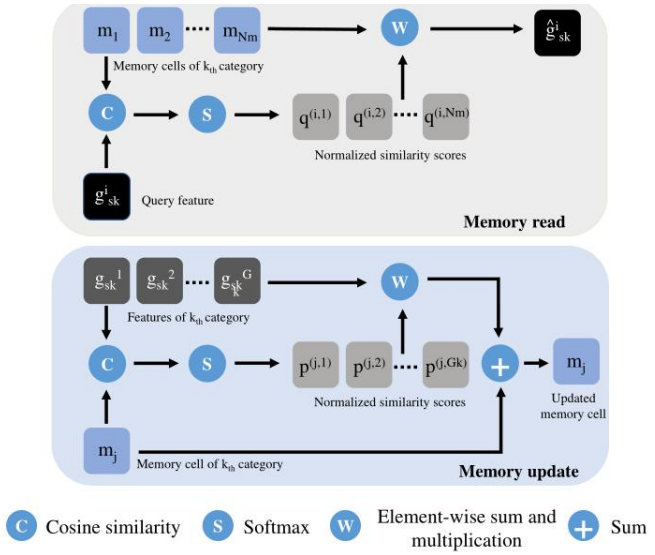

Figure 3. Read and write operations for the memory module.

记忆写：为了更新内存元素，我们只考虑源域图像，因为我们可以访问边界框标签来定位提取的特征映射中特定类别的特征。为了简单起见，我们来表示 $G_k = \{g^i_{s_k} \in \mathbb{R}^{1\times C}\}^{N_{s_k}}_{i=1}$ 所有属于 第 K 类的特征都在特征映射表 F_s 中。另外，记忆模块 M_k 中的每个记忆元表示为 $m_j \in \mathbb{R}^{1\times C}$，其中 $j \in \{1, ..., N_m\}$。首先，我们计算记忆元素 M_k 和表示第 K 类的特征 G_k 之间的规范化相似度量，如下:

$$p^{(j,i)} = \frac{\exp\left(m_j \cdot g^i_{s_k}\right)}{\sum_{l\in G_k} \exp\left(m_j \cdot g^l_{s_k}\right)},$$

其中，p 是一个 $N_m \times N_k$ 的相似性地图。我们利用记忆元素和类别特征之间的相似性，使用以下公式更新每个记忆元素:

$$m^j \leftarrow m^j + \sum_{i\in G_k} p^{(i,j)} g^i_{s_k}.$$

另外，请注意，如果源图像中没有第 K 类，我们不会更新相应内存模块的元素。在[34]之后，我们进一步规范了特征，确保内存元素不会离原始特征太远。这种正则化可以鼓励内存模块的紧凑性，从而减少类内的变化。这种损失是以 l2 距离惩罚的形式表示的:

$$\mathcal{L}_{cmp} = \sum_{j=1}^{N_m} \left\| m_j - g^p_{s_k} \right\|_2$$

其中，$g^p_{s_k}$ 是 m_j 的一个函数，表示 G_k 中和记忆元素 m_j 最相似的特性。除了使内存变得更加紧凑之外，我们还加入了一个唯一性约束，以减少内存元素的冗余。在[34]之后，我们利用记忆元素的三元损失，使得记忆模块中的每个记忆元素代表了基本类别的唯一原型。这一损失可以表述如下:

$$\mathcal{L}_{unq} = \sum_{j=1}^{N_m} max(\left\| m_j - g^p_{s_k} \right\|_2 - \left\| m_j - g^n_{s_k} \right\|_2, \alpha), \quad (6)$$

其中，α 表示三联体的损失边际，$g^p_{s_k}$ and $g^n_{s_k}$ 分别表示 G_k 中最相似和第二类似的特性。给出这些约束，内存的总体损失可以定义为:

$$\mathcal{L}_{mem} = \mathcal{L}_{cmp} + \mathcal{L}_{unq}. \quad (7)$$

记忆读：为了检索最相似的内存元素，我们计算内存 M_k 中每个项目与给定查询特性之间的相似度。请注意，查询特性可以来自源域，也可以来自目标域图像，即 $g^i_{s_k}$ or $g^i_{t_k}$。对于 $g^i_{s_k}$ 归一化相似度是用以下公式计算的:

$$q^{(i,j)} = \frac{\exp\left(m_j \cdot g^i_{s_k}\right)}{\sum_{l \in N_m} \exp\left(m_l \cdot g^i_{s_k}\right)}. \qquad (8)$$

给定这个归一化相似度 q，反演的特征 $\hat{g}^i_{s_k} \in \mathbb{R}^{1\times C}$ 可以表示如下:

$$\hat{g}^i_{s_k} = \sum_{j \in N_m} q^{(i,j)} m_j. \qquad (9)$$

另外，请注意，我们使用相同的公式从内存中读取源域和目标域特性。

3.3.2 注意机制

我们利用所有的内存模块，获得分类鉴别器的注意力地图。具体来说，为了计算目标特征映射 F_t，的注意力地图，我们查询每个元素 $f_t \in \mathbb{R}^{1\times C}$，到第 K 个记忆力模块 M_k，获得一个检索到的特征映射 $\hat{f}_t \in \mathbb{R}^{1\times C}$ 得到一张重新获取的特征图 $\hat{F}^k_t \in \mathbb{R}^{C\times H\times W}$，我们计算提取的特征映射 F_t 与检索到的特征映射 $\hat{F}^k_t$ 之间的元素相似度，来获得第 k 个类别感知鉴别器 $\sigma(F_t)_k$ 的注意映射，我们探索了两种相似度函数的选择来获得注意映射。

余弦相似度

文献中最常用的相似函数是余弦相似度。我们计算元素的余弦相似度得到 H*W 大小的 $\sigma(F_t)_k$：

$$\sigma(F_t)^{(h,\,w)}_k = \frac{F^{(h,w)}_t (\hat{F}^{k\ (h,w)}_t)^T}{||F^{(h,w)}_t||_2 ||\hat{F}^{k\ (h,w)}_t||_2}, \qquad (10)$$

Learned similarity.

虽然使用余弦相似度来计算注意图的准确性得到了合理的提高，但仔细观察这些图(见图 6 顶行)会发现，使用余弦相似度产生的注意并不准确。为了克服这个问题，我们探索了一个相似度度量，它是参数化的神经网络，可以在训练中学习。在这种情况下，我们使用一种度量学习方法，其中 F_t 和 $\hat{F}^k_t$ 首先分别通过一个网络 Θ_t 和 Θ^k_t。为了监督网络，我们利用了源数据集中的边界框信息。特别地，对于存在 k 类别的位置，我们将最大化 $\Theta_t(F_t)^{(h,w)}$ 和 $\Theta^k_t(\hat{F}^k_t)^{(h,w)}$ 的余弦相似度。以及在缺少相应类别的情况下最小化相似性（如图 4 所示）。然后，注意力图可以表示为：

$$\sigma(F_t)_k = Sim(\Theta_t(F_t), \Theta_t^k(\hat{F}_t^k)), \qquad (11)$$

式中，Sim（x，y）表示尺寸为 C×H×W 的张量 x 和 y 之间的元素余弦相似性，类似于等式 10。由此产生的注意力 σ（ •）k 的大小为 H×W。我们计算源图像和目标图像所有 K 类别的注意力。在将注意力转发到后续鉴别器之前，我们使用阈值对其进行二值化，如 0.5。如果标准化相似性大于 0.5 我们为地图指定 1，反之亦然。

Figure 6. 比较使用余弦相似性（顶行）和基于学习相似性的注意（底行）计算的注意图。虽然基于余弦相似度的分类方法对类别特征有合理的关注，但学习到的相似度得到了更准确的关注。

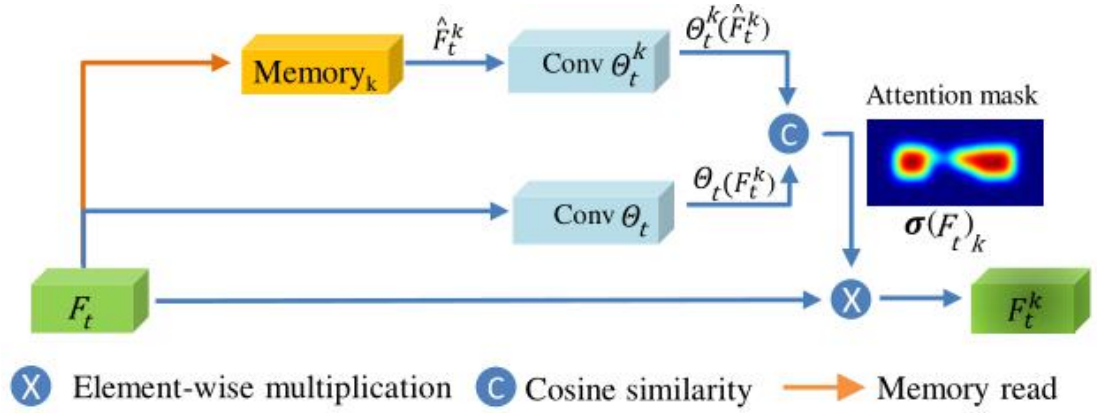

Figure 4. 记忆引导注意（MeGA）机制的前馈路径。输入源/目标特征映射被查询到第 k 个类别存储模块。通过读取操作，检索最接近的匹配元素，并通过学习到的相似性来预测注意图。注意图用于将 k 类信息发送至第 k 类别鉴别器模块。

3.4.MeGA CDA 的总体训练目标

对于我们的最终模型训练，我们在源域数据上添加有监督的检测损失，该数据具有图像和相应的带有类别标签的 bonding box 注释，如第 3.节所述。我们将监督检测损失表示为$\mathcal{L}_{det}$，包括[38]中描述的边界框回归损失和分类损失。包括前几节所述的全局、类别和记忆损失，建议方法的总体培训目标可以表示为：

$$\mathcal{L}_{cda}^{mega} = \mathcal{L}_{det}(X_s, b_s, y_s) + \beta \mathcal{L}_{gda}(X_s, X_t) + \gamma \sum_{k=1}^{K} \mathcal{L}_{cda}^{k}(X_s, X_t) + \lambda \mathcal{L}_{mem}, \quad (12)$$

4.实验与结果

4.1.实施细节

我们采用具有 VGG16 主干的 Faster-RCNN [38]网络，并使用学习率为 0.002,动量为 0.9 的 SGD 优化器进行训练。6 个 epoch 后学习率衰减为为 0.0002。全局和分类鉴别器由四个卷积层组成，具有 ReLU 非线性激活。批大小设置为 2，每个批包含一个来自源域的图像和一个来自目标域的图像。我们为每个类别使用 20 个内存项，每个内存项的维度为 1×1×C，其中 C 表示相应特征图中的通道数。网络 Θ_t, Θ_t^k 也由 4 个具有 ReLU 非线性的卷积层组成。我们对网络进行 10 个 epochs 的训练，并报告阈值为 0.5 的平均精度（mAP）。记忆损失的权重 λ 和域适配器的权重 β 、 γ 根据经验设置为 0.1 和 0.01。

4.2.定量比较

在本节中，我们将在三大适应类别（i）恶劣天气，（ii）合成到真实适应，以及（iii）交叉摄像机适应下，将所提出方法的性能与最近最先进的方法进行比较。

4.2.1 恶劣天气条件

在不同天气条件下稳定的目标检测性能对于安全关键应用（如自动驾驶汽车）至关重要。天气条件会引入图像伪影，这会对检测性能产生负面影响。为了评估该方法在恶劣天气下的有效性，我们分别使用雾城市景观和城市景观作为目标域和源域。

数据集：城市景观数据集[7]是在晴朗的天气条件下收集的，雾蒙蒙的城市景观[42]是通过在城市景观图像的顶部模拟薄雾创建的。城市景观和雾蒙蒙的城市景观都有 2975 张培训图像和 500 张验证图像，包括 8 个对象类别：人、骑手、汽车、卡车、公共汽车、火车、摩托车和自行车。

结果：在表 1 中，我们报告了我们的框架 MeGA-CDA 的性能，并与最近的自适应目标检测方法进行了比较。正如可以观察到的那样，MeGA-CDA 比现有方法有相当大的优势，同时比最近的最佳方法平均（绝对）mAP 提高了 2.5%. 此外，所提出的方法在所有类别中都表现良好，证明了将类别对齐与特征的全局对齐结合在一起的好处。

Table 1. Quantitative results (mAP) for Cityscapes → Foggy-Cityscapes dataset.

Method	person	rider	car	truck	bus	train	mcycle	bicycle	mAP
Source Only	25.8	33.7	35.2	13.0	28.2	9.1	18.7	31.4	24.4
DAFaster [6]	25.0	31.0	40.5	22.1	35.3	20.2	20.0	27.1	27.6
Strong-Weak [40]	29.9	42.3	43.5	24.5	36.2	32.6	30.0	35.3	34.3
MAF [17]	28.2	39.5	43.9	23.8	39.9	33.3	29.2	33.9	34.0
D&Match [25]	30.8	40.5	44.3	27.2	38.4	34.5	28.4	32.2	34.6
Selective DA [55]	33.5	38.0	48.5	26.5	39.0	23.3	28.0	33.6	33.8
MTOR [5]	30.6	41.4	44.0	21.9	38.6	40.6	28.3	35.6	35.1
ICR-CCR [52]	32.9	43.8	49.2	27.2	45.1	36.4	30.3	34.6	37.4
ATF [18]	34.6	47.0	50.0	23.7	43.3	38.7	33.4	38.8	38.7
MCAR [53]	32.0	42.1	43.9	**31.3**	44.1	43.4	**37.4**	36.6	38.8
Prior DA [46]	36.4	47.3	51.7	22.8	47.6	34.1	36.0	38.7	39.3
MeGA-CDA (ours)	**37.7**	**49.0**	**52.4**	25.4	**49.2**	**46.9**	34.5	**39.0**	**41.8**
Oracle [38]	37.2	48.2	52.7	35.2	52.2	48.5	35.3	38.8	43.5

4.2.2 合成数据自适应

合成数据为真实数据收集提供了一种廉价的替代方案，因为它更容易收集，并且通过适当的工程，合成数据可以自动注释。尽管计算机图形学有了长足的进步，但使用最先进的渲染引擎生成的照片真实感合成数据仍存在细微的图像瑕疵，这可能会导致在真实世界数据上的次优性能。

数据集：在本实验中，Sim10k[22]被用作源域，Cityscapes 被用作目标域。Sim10k 有 10000 幅图像，其中包含 58701 个类别边界框，由游戏引擎和防盗汽车渲染。我们使用所有 Sim10k 图像进行训练，并在 500 幅城市景观验证集的类别边界框上进行评估。

结果：在表 2 中，我们报告了以 Sim10K 合成数据为源，以城市景观为目标的框架图。所提出的方法 MeGA-CDA 将最近的最佳方法改进了 1.8%mAP（绝对改善）。由于我们正在从合成场景调整到真实场景，因此在调整第三个 conv 层的特性时，我们观察到更好的对齐。考虑到这个实验只有一类对象，所提出的类别感知对齐所取得的改进表明，记忆引导的注意确保了在正面和负面（背景）对象类别中更好地对齐。

Table 2. Quantitative results (mAP) for Sim10K → Cityscapes.

Method	mAP
Source Only	34.3
DAFaster [6]	38.9
MAF [17]	41.1
Strong-Weak [40]	40.1
ATF [18]	42.8
Selective DA [55]	43.0
MeGA-CDA (ours)	**44.8**
Oracle [38]	62.7

4.2.3 跨摄像机自适应

相机的内在和外在特性（如分辨率、失真、方向和位置）的差异会导致图像在质量、比例和视角方面捕捉到彼此不同的对象。虽然收集的数据可能是真实的，但这些域差异可能会导致严重的性能下降。

数据集：为了研究跨摄像机域间隙的这种影响，我们进行了两个 adapataion 实验，涉及

KITTI[13]和 Cityscapes 数据集。在第一个实验中，我们从 KITTI 适应到城市景观，在第二个实验中，我们从城市景观适应到 KITTI。请注意，KITTI 数据集由 7481 张图像组成，

结果：这两个实验的结果如表 3 所示。在这两个实验中，提出的方法能够取得相当大的改善，比最近最好的方法。从这些结果中，我们可以观察到，所提出的记忆引导的类别对齐在跨不同相机视图和光学特性桥接域间隙方面是有效的。

Table 3. Quantitative results (mAP) for KITTI → Cityscapes and Cityscapes → KITTI datasets.

Method	KITTI → City	City → KITTI
Source Only	30.2	53.5
DAFaster [6]	38.5	64.1
MAF [17]	41.0	72.1
Strong-Weak [40]	37.9	71.0
Selective DA [55]	42.5	-
ATF [18]	42.1	73.5
MeGA-CDA (ours)	**43.0**	**75.5**

4.3.消融研究

我们通过迭代添加每个模块来研究所提出方法的不同组件 MeGA-CDA 的影响。我们使用城市景观→雾城市景观适应实验来实现消融研究。

定量分析：通过使用全局域鉴别器(GDA)调整 conv5 特征，我们观察到相对于仅源基线的合理改进。通过使用基于余弦相似性的注意训练（cosine similarity-based Attention）的记忆引导类（memory-guided）别鉴别器（GDA+CDA+MA）扩充全局域鉴别器，我们得到了 0.6%mAP 的进一步改进。这说明了在域自适应期间添加类别信息的好处。当在多层（conv4 和 conv5）上应用 GDA-CDA-MA 时，我们观察到进一步改善约 3%。最后，我们证明了使用度量学习相似性（LS）方法（MeGA-CDA）加强内存模块可以增强记忆库捕获数据特征的能力，并导致进一步的改进。具体来说，当 MeGA-CDA 应用于 conv5 时，我们观察到与具有余弦相似性的 GDA CDA MA 基线相比改善 2.1%。此外，在 conv4 和 conv5 块上应用 MeGA CDA 可额外提高 2%。如前所述，该子网络使用来自源域的 ground truth 边界框的弱监督进行训练，因此，它不需要任何附加注释。

定性分析：我们比较了图 5 中针对城市景观→雾城市景观适应实验的全局对齐方法和拟议类别对齐方法的检测结果。如图 5 所示，基于全局对齐的方法会导致错误，例如漏检（假阴性）或假阳性。例如，背景被检测为对象（底行），或者对象被错误地指定了错误的类别和边界框大小（顶行）。最可能的原因是特征的负迁移，因为全局适应以类别不可知的方式对齐特征。在这两种情况下，建议的分类对齐能够通过更好地对抗特征的负迁移来纠正错误。在图 6 中，我们展示了在 MeGA CDA 培训期间为汽车类别生成的注意力地图。为了可视化，我们将内存模块生成的注意力图覆盖在图像上。第一行和第二行分别显示使用余弦相似度和基于度量学习的相似度计算的注意图。可以观察到，基于余弦相似性的注意提供了对汽车类别位置的合理关注。然而，通过学习相似度，我们在注意力跨越汽车大部分区域时取得了更大的效率。这是预期的，因为学习到的相似性是通过度量学习进行训练的，并且来自源域 ground truth 的弱监督，从而导致记忆项的引导学习。

图 5。定性检测结果。全局对齐会导致漏检。相比之下，所提出的方法在实现高质量检测的同时减少了误报。

5. Conclusions

我们提出了一种基于类别感知的特征对齐方法，用于域自适应目标检测。具体来说，我们通过引入类别感知鉴别器将类别信息纳入域对齐过程。为了克服缺乏类别标签的问题，特别是在目标领域，我们提出了记忆引导的注意机制，该机制生成类别特定的注意映射，用于将特征路由到适当的类别特定鉴别器中。通过这样做，我们能够缓解负迁移问题，从而实现更好的整体协调。MeGA CDA 在多个基准数据集上进行了评估，结果表明，其性能远远优于现有方法。