

AdvPull：一种针对 RGB-红外行人重识别的新型物理对抗攻击方法

吴蕊寒

班级-学号 2218039-22009101379

摘要 行人重识别（Re-ID）是一项在不同摄像机间匹配行人图像的任务，广泛应用于监控和安全领域。近年来，许多研究致力于将夜间红外图像中捕获的目标（如犯罪嫌疑人）与白天 RGB 图像中的行人进行匹配，以追踪其行踪，这种任务被称为 RGB-红外行人重识别（RGB-IR Re-ID）。然而，基于深度神经网络（DNN）的 RGB-IR Re-ID 模型被证明存在漏洞，容易受到攻击，这引发了安全方面的担忧。

本文首次提出了一种针对 RGB-IR Re-ID 的物理对抗攻击方法。我们通过一种名为 AdvPull 的新算法，在气凝胶片上生成对抗纹理，并将其附着在衣物上，使得红外和 RGB 摄像机下捕获的行人图像特征差异显著，无法被 Re-ID 模型匹配。这是一种黑盒攻击方法，具有很高的迁移性。实验结果表明，AdvPull 方法在 SYSU-MM01 数据集上，与未受攻击时相比，Rank-1 识别率降低了 99.99%，mAP 降低了 96.2%；在 Regdb 数据集上，Rank-1 识别率降低了 99.99%，mAP 降低了 90.4%。这表明该方法能够有效降低 RGB-IR Re-ID 系统的识别精度，为保护个人隐私和增强模型鲁棒性提供了一种新的解决方案。未来，我们将进一步优化该方法，并探索其在其他载体中的应用潜力。

关键词 行人重识别；对抗攻击；RGB-IR；隐私保护；模型鲁棒性

AdvPull: A Novel Physical Adversarial Attack for RGB-Infrared Person Re-Identification

Ruihan Wu¹⁾ Yue Su²⁾ Haonan Shi²⁾

¹⁾School of Cyber Engineering, Xidian University ²⁾School of Artificial Intelligence, Xidian University

Abstract Pedestrian re-identification (Re-ID) is a task that matches pedestrian images across different cameras and is widely used in surveillance and security applications. In recent years, many Re-ID efforts have focused on matching targets (such as criminal suspects) captured in infrared images at night with pedestrians captured in RGB images during the day to track their movements. This type of work is known as RGB-infrared pedestrian re-identification (RGB-IR Re-ID). However, RGB-IR Re-ID models based on deep neural networks (DNNs) have been proven to be vulnerable and susceptible to attacks, raising significant security concerns.

In this paper, we present a novel physical adversarial attack method for RGB-IR Re-ID. Using an algorithm called AdvPull, we generate adversarial textures on aerogel sheets and attach these sheets to clothing. This causes the features of pedestrian images captured by IR and RGB cameras to diverge significantly, making them unmatchable by Re-ID models. This method is a black-box attack and highly transferable. Experimental results show that compared to the unattacked state, the Rank-1 recognition rate on the SYSU-MM01 dataset decreased by 99.99%，and the mAP decreased by 96.2%. On the Regdb dataset, the Rank-1 recognition rate decreased by 99.99%，and the mAP decreased by 90.4%. These results demonstrate that our method can effectively reduce the recognition accuracy of RGB-IR Re-ID systems, offering a new solution for protecting individual privacy and enhancing model

robustness. Future work will focus on further optimizing this method and exploring its potential applications in other fields.

Key words Person Re-Identification, Adversarial Attack, RGB-IR, Privacy Protection, Model Robustness

1 引言

1.1 作品产生背景

行人重识别(Re-ID)是一项在由大量非重叠多视角摄像机组成的网络中,从海量视频帧中查找目标行人的任务,可被视为一种目标检索问题。传统的行人重识别主要基于RGB图像,然而在夜间,摄像机很难捕捉到清晰的RGB图像,此时通常会采用红外摄影。例如,在追捕犯罪嫌疑人的任务中,许多犯罪分子在夜间作案,我们往往只能获取其夜间行动的红外图像;此时,若想重新定位犯罪嫌疑人的行踪,就必须调用其他场景的摄像机拍摄的RGB图像和夜间拍摄的红外图像,通过相似度评分排序来锁定目标人物。反之,当我们的查询目标人物在RGB场景中被拍摄时,我们也可以补充使用红外场景来查询其在红外场景中的匹配。这种跨模态的行人重识别被称为RGB-红外行人重识别(RGB-IR Re-ID)。

近年来,受深度学习的影响,许多行人重识别工作都基于深度神经网络(DNN)展开,但研究表明DNN非常容易受到攻击,这给行人重识别工作带来了巨大的安全隐患:模型中的漏洞往往意味着犯罪嫌疑人能够成功逃脱,以及其他诸多问题。近年来,针对深度Re-ID模型的对抗攻击层出不穷,这些攻击被证明具有以下含义:(1)从监控方的角度来看,对抗攻击有助于发现模型中的漏洞。监控方可以引入对抗样本对模型进行训练,使模型更具鲁棒性,以应对紧急情况。(2)从被监控方的角度来看,某些名人需要保护自己的隐私,防止不良分子通过入侵监控系统重新识别他们并获取其行踪。因此,他们需要将图像处理成对抗样本,以防止被重新识别。

然而,现有的攻击方法在面对实际情况时存在诸多问题:(1)除了一篇关于物理对抗攻击的文章外,大多数现有的Re-ID攻击研究都在数字领域实施。其大多数攻击方法是对图像本身的像素进行扰动。实际上,被监控对象很难获得捕获图像信息的权限,这意味着我们对图像本身的扰动攻击方法在实际意义上往往无效。然而,物理对抗攻击,例

如在行人的衣服上设计对抗图案,可以确保我们的对抗样本出现在任何捕获的图像中。同时,这种攻击模式可以在黑盒中实施,因此也适应了我们无法提前了解Re-ID模型信息的要求。另一方面,被监控对象通常不知道自己何时何地被哪种摄像机模式监控。这也要求我们提前准备好对抗样本,而不是实时处理图像。这只有通过物理对抗攻击才能实现。(2)现有的攻击都针对RGB单模态Re-ID任务,这意味着当我们使用红外图像作为补充来匹配目标时,这种攻击会失败;另一方面,当红外图像本身被用作查询目标时,针对它的攻击也是一个未被充分研究的领域。在这种情况下,掌握RGB-IR Re-ID跨模态攻击的犯罪嫌疑人可以轻松逃脱视频监控的追捕。

然而,现有的可用于参考的跨模态攻击方法基于识别和分类领域。这些方法相对简单,只是将红外和RGB模式中的分类错误结合起来设计一个二元损失。无论是分类还是重识别任务,许多模型都可以学习融合两种模态的特征来高效完成任务,这要求我们更全面地审视损失。

1.2 基本贡献及应用价值

为了解决上述问题,我们通过AdvPull方法设计了一种物理对抗模式,即带有对抗纹理的气凝胶贴片。我们为现有的跨模态Re-ID引入了特征融合损失、相似度排序损失和模态转换损失;我们尽力扩大红外图像和RGB图像中行人之间的特征差距,并减小两种模态中行人之间的差异,以实现攻击效果;同时我们将对抗性攻击转化为优化问题,最终确定最优气凝胶贴片的形状、温度和纹理像素。本文的主要贡献如下:

1、在针对Re-ID的攻击中,我们首次考虑了跨模态攻击,且这种攻击是主动的、黑盒的。除此之外,我们的方法是少数可以在物理世界中实施的攻击之一。

2、整个RGB-红外对抗攻击领域,与以往分别攻击两种模态的方法相比,我们首次考虑了模态间的信息传递和特征融合,并设计了新颖的攻击方法,可以为后来者提供启发。

3、本方法在SYSU-MM01数据集上,将AGW

模型的 Rank-1 识别率降至 0.0%，mAP 降至 3.8%；在 Regdb 数据集上，将 AGW 模型的 Rank-1 识别率降至 0.01%，mAP 降至 4.8%。这些关键指标的显著下降，充分证明了我们的攻击方法对主流 Re-ID 模型具有强大的攻击效果和广泛的适用性。

2 技术方案

2.1 特征提取

考虑到现有的跨模态再识别模型大多使用神经网络提取 RGB 图像红外图像的共同特征来判断它们是否匹配。我们力求在引入对抗补丁后，拉长两者特征之间的距离，从而使 Re-ID 无法正确匹配目标。假设目标人物的 RGB 图像为 I_{rgb} ，红外图像为 I_{ir} ，而我们引入的对抗补丁为 δ ，因此我们认为引入对抗补丁后的 RGB 图像为 $I'_{rgb} = I_{rgb} + \delta$, $I'_{ir} = I_{ir} + \delta$ ，这些都是我们的对抗样本。需要注意的是， δ 在 RGB 图像和红外图像中的表现不同。在 RGB 图像中， δ 上的对抗贴片表面的纹理对图像像素的影响起主要作用，而在红外图像中，主要是 δ 补丁的材料通过自身温度释放热辐射，改变红外图像中的像素。

首先使用 FAST 算法提取 I'_{rgb} 的特征点，将特征点集合排序为 $\{p_r^1, p_r^2 \dots p_r^n\}$ 。同样，提取 I'_{ir} 的特征点排序为 $\{p_r^1, p_r^2 \dots p_r^m\}$ 。其后，使用 rRBRIEF 算法^[1,2]将这些特征点表示为 k 阶二元描述符。

$$\begin{aligned} p_r^t &= (a_1, a_2 \dots a_k), \quad t = 1, 2 \dots n \\ p_r^j &= (a_1, a_2 \dots a_k), \quad j = 1, 2 \dots m \end{aligned}$$

其中， $a_1, a_2 \dots a_k$ 取 0 或 1。

使用汉明距离 $d(x, y)$ 来表示特征之间的距离，表示为： $d(p_r^t, p_r^j) = \sum_{s=1}^k |p_r^t[s] - p_r^j[s]|$ 。目标最大化匹配特征点之间的距离，因此特征匹配的损失设计为：

$$\begin{aligned} L_{fea} &= - \sum_{j=1}^m d(p_r^j, \arg \min_{p_r^t} d(p_r^t, p_r^j)) \quad t \in \mathbb{Z}, 1 \\ &\leq t \leq n \end{aligned}$$

2.2 学习误差排序

Re-ID 问题归根结底是一个相似度排序问题，只要能够破坏模型的排序，尽可能降低正确匹配图像的相似度排序，也就是让模型学习错误的排序，就能实现有效攻击。假设攻击的深度 Re-ID 模型是 $f_\theta(\cdot, \cdot)$ ， θ 为模型参数。由于本攻击是先发制人

的黑盒攻击，我们无法提前知道模型参数。不过，由于大多数深度 Re-ID 模型的底层相似性评估方法仍然基于图像之间的余弦相似性，因此可以将 $f_\theta(\cdot, \cdot)$ 表示为：

$$f_\theta(I'_{rgb}, I'_{ir}) = \frac{\sum_{s=1}^n p_r^s p_i^s}{\|I'_{rgb}\|_2 \cdot \|I'_{ir}\|_2}$$

其中， p_r^s 和 p_i^s 分别代表 RGB 图像和红外图像中的第 s 个像素点。使用 $rank(f_\theta(I'_{rgb}, I'_{ir}))$ 表示这两张图片在图库中的相似度排序。我们认为排名 k 的图片对最终被识别为匹配对象的概率服从参数为 λ 的泊松分布，即 $P(rank = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ ，目标表示为：

$$\min \sum_{i=1}^I \sum_{k=1}^K P(rank(f_\theta(I'_{ir}[k], I'_{rgb}[k])))$$

这是在物体探测器处于红外模式下产出的公式，而我们希望在可见光视频库中完成重新识别任务建立。这里 i 代表第 i 个摄像机拍摄的视频图库。而 k 表示目标行人的 ID 是 k 。但正如 Wang 等人^[3]的文献中所述，单纯使用图像攻击是无效的，因为我们的对抗样本很可能无法匹配其他 ID，并且过于突兀容易被识别为异常样本。如果我们的对抗样本可以与其他 ID 匹配，达到一定程度的冒充，那么我们的攻击就可以更加隐蔽：

$$\max \sum_{i=1}^I \sum_{k=1}^K \sum_{t=1, t \neq K}^T P(rank(f_\theta(I'_{ir}[k], I'_{rgb-i}[t])))$$

这里的 t 代表与 k 不同的行人 ID。通过最大化不同 ID 之间的匹配概率，我们就能达到冒名顶替的效果。最后，我们的损失 h 函数可以描述为：

$$\begin{aligned} L_{er} &= \sum_{i=1}^I \sum_{k=1}^K P(rank(f_\theta(I'_{ir}[k], I'_{rgb}[k]))) \\ &- \sum_{i=1}^I \sum_{k=1}^K \sum_{t=1, t \neq K}^T P(rank(f_\theta(I'_{ir}[k], I'_{rgb-i}[t]))) \end{aligned}$$

同理的，如果我们使用 RGB 图像作为查询探针，红外图像作为匹配探针，损失可以描述为：

$$L_{er} = \sum_{i=1}^I \sum_{k=1}^K P(rank(f_\theta(I'_{rgb}[k], I'_{ir-i}[k])))$$

$$-\sum_{i=1}^I \sum_{k=1}^K \sum_{t=1, t \neq K}^T P\left(\text{rank}\left(f_\theta\left(I'_{rgb}[k], I'_{ir-i}[t]\right)\right)\right)$$

2.3 模式间的信息消融

模态间信息消减是为了防止 Re-ID 系统通过模态转换将 RGB 图像中恢复的灰度信息与红外图像进行匹配，信息消减本身也可以减少两种模态的特征重叠。由于现有的灰度还原方法大多基于三个通道的加权平均或其中一个通道的一致组合，因此我们不妨提取 RGB 图像中的每个通道信息，并在红外图像中构建信息损失。即将 RGB 图像分解为：

$I'_R = (I'_{rgb_x}, I'_{rgb_y}, I'_{rgb_z})$, $I'_G = (I'_{rgb_x}, I'_{rgb_y}, I'_{rgb_z})$,
 $I'_B = (I'_{rgb_x}, I'_{rgb_y}, I'_{rgb_z})$, 这是为了增强每幅 RGB 图像的单通道，并将其分解为三幅灰度图像。这里用 x 表示 RGB 图像中某一点的灰度值，我们将三通道增强图像和红外图像的灰度分布离散函数分别表示为 $I'_R(x)$, $I'_G(x)$, $I'_B(x)$, $I'_{ir}(y)$ 。使用 $P_R(x, y)$, $P_G(x, y)$, $P_B(x, y)$ 来表示同一坐标像素点上信道增强图和红外图像的联合概率分布函数。所以它们之间的归一化互信息可分别表示为：

$$\begin{aligned} NMI(I'_R, I'_{ir}) \\ = \frac{\sum_{x \in X} I'_R(x) \log I'_R(x) + \sum_{y \in Y} I'_{ir}(y) \log I'_{ir}(y)}{\sum_{x \in X} \sum_{y \in Y} P_R(x, y) \log \frac{P_R(x, y)}{I'_R(x) I'_{ir}(y)}} \end{aligned}$$

$$\begin{aligned} NMI(I'_G, I'_{ir}) \\ = \frac{\sum_{x \in X} I'_G(x) \log I'_G(x) + \sum_{y \in Y} I'_{ir}(y) \log I'_{ir}(y)}{\sum_{x \in X} \sum_{y \in Y} P_G(x, y) \log \frac{P_G(x, y)}{I'_G(x) I'_{ir}(y)}} \end{aligned}$$

$$\begin{aligned} NMI(I'_B, I'_{ir}) \\ = \frac{\sum_{x \in X} I'_B(x) \log I'_B(x) + \sum_{y \in Y} I'_{ir}(y) \log I'_{ir}(y)}{\sum_{x \in X} \sum_{y \in Y} P_B(x, y) \log \frac{P_B(x, y)}{I'_B(x) I'_{ir}(y)}} \end{aligned}$$

目标是使归一化后互信息尽可能小，以降低模式相互切换的可能性和信息传递的强度，同时在一定程度上降低模式之间的相似性，损失函数设计为：

$$\begin{aligned} L_{inf} = - & (NMI(I'_R, I'_{ir}) + NMI(I'_G, I'_{ir}) \\ & + NMI(I'_B, I'_{ir})) \end{aligned}$$

最终目标函数如下：

$$\min L_{adv} = L_{fea} + \lambda_1 L_{er} + \lambda_2 L_{inf}$$

上述 λ_1 和 λ_2 是用于控制权重的系数。

2.4 对抗样本的生成

对抗样本的生成本质上是生成对抗补丁 δ 的过程。但是在应用优化方法生成 δ 之前，我们需要考虑两个问题：补丁的隐蔽性和物理变化场景中的通用性。

2.4.1 补丁隐藏

虽然之前的目标函数可以全面攻克深度跨模态再识别模型，但如果这个目标函数优化生成的补丁 δ 不受到特定的约束，在现实中生成的补丁很有可能分布怪异，容易引起注意。这样做的后果是，如果现实场景中有检查人员，会立即发现异常。为此，我们必须对生成的 δ 做一定的限制，使其生成的纹理更加自然，符合正常人的审美观。首先参照 shaferi et al.^[4] 的观点来最小化总的波动函数 $TV(\delta)$ 。该函数可以通过限制相邻像素的变化来实现整体像素分布的均匀性和平滑性。 $\min TV(\delta) = \sum_{p,q} ((\delta_{p,q} - \delta_{p+1,q})^2 + (\delta_{p,q} - \delta_{p,q+1})^2)^{\frac{1}{2}}$ 这样设计出的补丁显然更符合人体的视觉常态。

另一方面，考虑到我们的补丁最终要贴在衣服上，因此我们努力使其看起来与普通衣服无异。针对这一目标，我们提出了创新的解决方案。我们将贴片的形状限定为 S ，然后寻找与其大小最接近的卡通图案形状 S' （如熊、猫等），并通过边界填充将 S 映射到 S' （具体来说，我们使用了外部 WRAP 方法）。 $patch = \delta(S) + \text{WRAP}_\delta(S' - S) \text{ s.t. } S \subseteq S'$

这样一来，装载对抗样本的人物所穿的衣服与街上的普通行人衣物无异。

2.4.2 物理变化场景的通用性

客观上来说，当我们把在数字领域生成的对抗性补丁移植到现实世界时，会出现很多问题。从摄像机的角度来看，行人之间的距离、当天天气影响的光照以及摄像机拍照的角度都会影响对抗样本

的质量。至于行人本身，对抗补丁本身是贴在衣服上的，因此行走时衣服的褶皱也会影响对抗样本的效果。以上这些在现实中实际存在的情况，都对我们的对抗攻击方法的鲁棒性提出了挑战。

针对上述挑战，我们决定引入 EOT:Expectation Over Transformation 技术^[5]来增强对抗样本的鲁棒性。具体来说，这是一种使用 3D 渲染函数 $t(\cdot)$ 来模拟真实世界中光照、距离、角度和形变变化的模型。通过在优化过程中对对抗样本进行 EOT 处理，我们可以让它在面对真实物理场景时更加稳健。因此，我们可以这样更新对抗样本：

$$I'_{rgb} = t(I_{rgb} + \delta) \text{ s.t. } \mathbb{E}_{t \sim T} [d(I'_{rgb}, t(I_{rgb}))] \leq \epsilon$$

此外，它还限制了原始图像与经历 EOT 变化后的对抗图像之间的差异小于一定范围。这可以在一定程度上限制 EOT 的变化，更符合物理世界的客观规律。

3 实验方案

3.1 实验设置

为了验 AdvPull 方法的有效性，在两个公开的 RGB-IR 行人重识别数据集上进行了实验：SYSU-MM01 和 Regdb。我们使用 Rank-1（排名第一的图像匹配的概率），Rank-5（排名前五的图像匹配的概率），Rank-10（排名前十的图像匹配的概率），mAP（平均精度均值），ss（标准差）来进行评估。

实验环境：

实验在标准的深度学习环境中进行，使用 PyTorch 框架实现 AdvPull 算法。我们采用了 Adam 优化器，学习率设置为 0.001，并在 NVIDIA GeForce RTX 3090 GPU 上进行训练和测试。

3.2 实验结果

我们在 SYSU-MM01 和 Regdb 数据集上对多种主流的 Re-ID 模型进行了攻击实验，结果如表 1 所示。

实验结果表明，AdvPull 方法在 SYSU-MM01 和 Regdb 数据集上对多种主流的 Re-ID 模型均取得了显著的攻击效果。在 SYSU-MM01 数据集上，AdvPull 方法使 AGW、DDAG 和 DEEN 模型

的 Rank-1 识别率分别降至 0.0%、0.01% 和 0.0%，mAP 分别降至 3.8%、4.3% 和 3.9%；在 Regdb 数据集上，相应指标分别降至 0.01%、0.0% 和 0.0%（Rank-1），以及 4.8%、3.6% 和 4.5%（mAP）。这些显著下降的关键指标表明，AdvPull 方法对不同 Re-ID 模型具有强大的攻击效果和良好的迁移性。

表 1 攻击实验结果

模型	数据集	Rank-1	Rank-5	Rank-10	mAP	ss
AGW	SYSU-MM01	0.0%	0.0%	0.0%	3.8%	0.393
	Regdb	0.01%	0.0%	0.0%	4.8%	0.452
DDAG	SYSU-MM01	0.01%	0.02%	0.04%	4.3%	0.422
	Regdb	0.0%	0.0%	0.0%	3.6%	0.362
DEEN	SYSU-MM01	0.0%	0.0%	0.0%	3.9%	0.401
	Regdb	0.0%	0.0%	0.0%	4.5%	0.445

4 总结以及展望

本文引入物理对抗方案，增强了攻击的可操作性。考虑了物理场景的通用性，隐藏性相对较高。然而，本方法在实际应用中面临挑战，如对抗纹理的生成与固定需特定技术条件，且在复杂现实环境中，对抗样本的鲁棒性可能受限。未来工作将聚焦于提升对抗样本的鲁棒性和隐蔽性，以拓展其在更多实际场景中的应用。同时，我们还将探索该方法在其他行人重识别任务中的应用，并研究如何应对模型防御机制，以进一步提升 AdvPull 方法的攻击效果和实用性。

参 考 文 献

- [1] Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). BRIEF: Binary robust independent elementary features. In *Computer Vision-ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11* (pp. 778-792). Springer Berlin Heidelberg.
- [2] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011, November). ORB: An efficient alternative to SIFT or SURF. In *2011 International conference on computer vision* (pp. 2564-2571). IEEE.
- [3] Wang, Lin, et al. "Attack is the best defense: Towards preemptive-protection person re-identification." Proceedings of the 30th ACM international conference on multimedia. 2022.

- [4] Shafahi, Ali, et al. "Adversarial training for free!" *Advances in neural information processing systems* 32 (2019).
- [5] Athalye, Anish, et al. "Synthesizing robust adversarial examples." *International conference on machine learning*. PMLR, 2018.
- [6] Wang, Zhibo, et al. "advpattern: Physical-world attacks on deep person re-identification via adversarially transformable patterns." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.
- [7] Yang, Fengxiang, et al. "Learning to attack real-world models for person re-identification via virtual-guided meta-learning." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. No. 4. 2021.
- [8] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep Residual Learning for Image Recognition[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016: 770-778.
- [9] MA Xingjun, LI Bo, WANG Yisen, et al. Adversarial Attacks and Defenses in Deep Learning[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 32(8): 3254-3270.
- [10] LECUYER Mathias, ATLIDAKIS Vaggelis, GEAMBASU Roxana, et al. Adversarial Robustness Toolbox: A Library for Adversarial Machine Learning[EB/OL]. arXiv:1807.01069, 2018.
- [11] WANG Zhibo, LIU Zhenyu, SUN Yifan, et al. Adversarial Attacks on Deep Person Re-identification Models[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019: 3771-3780.
- [12] YANG Fengxiang, LIU Zhenyu, SUN Yifan, et al. Learning to Attack Real-World Models for Person Re-identification via Virtual-Guided Meta-Learning[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2021, 35(6): 5664-5672.