



(12) 发明专利

(10) 授权公告号 CN 118396830 B

(45) 授权公告日 2024. 08. 30

(21) 申请号 202410851622.9

(22) 申请日 2024.06.28

(65) 同一申请的已公布的文献号

申请公布号 CN 118396830 A

(43) 申请公布日 2024.07.26

(73) 专利权人 齐鲁工业大学(山东省科学院)

地址 250300 山东省济南市长清区大学路
3501号

(72) 发明人 夏之秋 田鹏飞 王长旭 魏子麒

马宾 王玉立 张昊 李琦 李健

咸永锦 王晓雨 韩冰 王春鹏

(74) 专利代理机构 北京中索知识产权代理有限公司

公司 11640

专利代理师 商金婷

(51) Int.Cl.

G06F 30/27 (2020.01)

(56) 对比文件

田鹏飞.“基于生成对抗网络的不可感知水印攻击算法研究”.《中国硕士学位论文全文数据库》.2024,全文.

审查员 巩玉

权利要求书2页 说明书6页 附图3页

(54) 发明名称

一种无非线性水印攻击系统

(57) 摘要

本发明公开一种无非线性水印攻击系统,涉及数字水印技术领域,包括以下步骤:采用经典的U-Net架构和跳跃连接,同时使用融入通道注意力的残差网络对RGB图像进行水印攻击任务;在非线性水印攻击网络、残差网络和通道注意力中均使用无非线性残差块,进行非线性的去除处理;无非线性残差块,接收输入特征图,并通过一系列的卷积操作提取特征,卷积层用于进行卷积操作;转置卷积用于进行上采样操作。本发明在水印攻击能力和水印攻击不可感知性两方面表现上均优越于现有的水印攻击方法。



1. 一种无非线性水印攻击系统,其特征在于,包括:

采用经典的U-Net架构和跳跃连接,同时使用融入通道注意力的残差网络对RGB图像进行水印攻击任务;

在非线性水印攻击网络、残差网络和通道注意力中均使用无非线性残差块,进行非线性的去除处理;

无非线性残差块,接收输入特征图,并通过一系列的卷积操作提取特征,卷积层用于进行卷积操作;转置卷积用于进行上采样操作,非线性水印攻击网络使用了这两个层来进行特征图的下采样和上采样;

采用GeLU,并且对GeLU进行非线性的去除,有效解决的训练过程中的梯度消失问题,增强了神经元的感知能力,提高了对输入数据的表达能力;

无非线性残差块内引入注意力机制,并且对通道注意力机制进行了进一步优化,提出一种简化的不含有非线性的通道注意力;

无非线性残差块使用门控线性单元进行非线性去除工作;

GLU是一种常用的神经网络模块,结合线性变换和门控函数,定义为:

$$GLU(X, f, g, \sigma) = f(X) \odot \sigma(g(X)) \quad (1);$$

其中: X 表示特征映射;

f 和 g 是线性变换器;

σ 是一个非线性激活函数

$\odot$ 表示逐元素乘法;

通道注意力首先对输入特征进行全局平均池化操作,随后将全局平均池化后的特征传递给两个线性层和ReLU,用于进行特征压缩和映射,最后将特征映射后的结果输入到Sigmoid函数中,得到通道注意力权重,过将计算得到的通道注意力权重与输入特征进行逐元素相乘,实现对输入特征的加权,使用公式表示为:

$$CA(X) = X * \sigma(W_2 \max(0, W_1 pool(X))) \quad (5);$$

其中: X 表示特征图;

$pool$ 表示将空间信息居合道通道中的全局平均池化操作;

σ 是一个非线性激活函数Sigmoid;

W_1 和 W_2 表示全连接层;

倘若将通道注意力机制看成一个函数,则式(5)简化为:

$$CA(X) = X * G(X) \quad (6);$$

将 $G(X)$ 看作 $\sigma(W_2 \max(0, W_1 pool(X)))$, 保留通道注意力中较为重要的部分,提出简化通道注意力机制具体公式如下所示:

$$SCA(X) = X * W pool(X) \quad (7);$$

非线性激活函数可表示为:

$$GELU(x) = x\phi(x) \quad (2);$$

其中: ϕ 表示标准正态分布的积累分布函数;

GELU可以通过以下方式进行近似实现:

$$0.5x \left(1 + \tanh \left[\sqrt{\frac{2}{\pi}} \left(x + 0.044715x^3 \right) \right] \right) \quad (3);$$

根据式(1)和式(2),发现GELU是GLU的一种特殊情况,即 f 和 g 是恒等函数,将 σ 看作 ϕ ,由于这种相似性的存在,将GLU视为激活函数用于水印攻击领域,并且GLU本身包含非线性,并不依赖于 σ ,即便移除 σ ,GLU依然不包含非线性,因此对GELU进行简化,提出无非线性GeLU;

直接在通道维度上将特征映射分成两部分并相乘,SimpleGate通过逐元素乘法来实现,具体公式如下所示:

$$SimpleGate(X, Y) = X \odot Y \quad (4);$$

其中: X 和 Y 是相同大小的特征图,通过将GELU简化为SimpleGate,水印攻击综合能力显著提升。

一种无非线性水印攻击系统

技术领域

[0001] 本发明涉及数字水印技术领域,具体地讲,涉及一种无非线性水印攻击系统。

背景技术

[0002] 数字水印技术是一种可以将隐藏的信息嵌入到数字媒体中的技术。其目的是为了保护数字媒体的版权和完整性,以防止非法复制和篡改,广泛应用于版权保护、身份验证、数字完整性检查等领域。数字水印技术和水印攻击技术有着密切的关系,水印攻击技术是指对数字媒体所嵌入的水印信息进行攻击的各种方法和技术。

[0003] 事物的发展总是辩证的,数字水印技术和水印攻击技术是相辅相成的,数字水印技术想要不断优化需要面对更有效的水印攻击技术的挑战。鲁棒的水印算法需要在不断的历练中成型,倘若没有有效的水印攻击算法对水印算法进行磨砺,将导致水印算法的发展停滞不前。因此发展水印攻击算法是来完善数字水印鲁棒性测试工具十分必要的。

发明内容

[0004] 本发明要解决的技术问题是提供一种无非线性水印攻击系统,在水印攻击能力和水印攻击不可感知性两方面表现上均优越于现有的水印攻击方法。

[0005] 本发明采用如下技术方案实现发明目的:

[0006] 一种无非线性水印攻击系统,其特征在于,包括以下步骤:

[0007] 采用经典的U-Net架构和跳跃连接,同时使用融入通道注意力的残差网络对RGB图像进行水印攻击任务;

[0008] 在非线性水印攻击网络、残差网络和通道注意力中均使用无非线性残差块,进行非线性的去除处理;

[0009] 无非线性残差块,接收输入特征图,并通过一系列的卷积操作提取特征,卷积层用于进行卷积操作;转置卷积用于进行上采样操作,非线性水印攻击网络使用了这两个层来进行特征图的下采样和上采样;

[0010] 采用GeLU,并且对GeLU进行非线性的去除,有效解决的训练过程中的梯度消失问题,增强了神经元的感知能力,提高了对输入数据的表达能力;

[0011] 无非线性残差块内引入注意力机制,并且对通道注意力机制进行了进一步优化,提出一种简化的不含有非线性的通道注意力。

[0012] 作为本技术方案的进一步限定,无非线性残差块使用门控线性单元进行非线性去除工作;

[0013] GLU是一种常用的神经网络模块,结合线性变换和门控函数,定义为:

[0014]
$$\text{GLU}(X, f, g, \sigma) = f(X) \odot \sigma(g(X)) \quad (1);$$

[0015] 其中: X 表示特征映射;

[0016] f 和 g 是线性变换器;

[0017] σ 是一个非线性激活函数

[0018] $\odot$ 表示逐元素乘法。

[0019] 作为本技术方案的进一步限定,非线性激活函数可表示为:

$$[0020] \quad GELU(x) = x\phi(x) \quad (2);$$

[0021] 其中: ϕ 表示标准正态分布的积累分布函数;

[0022] GELU可以通过以下方式进行近似实现:

$$[0023] \quad 0.5x \left(1 + \tanh \left[\sqrt{\frac{2}{\pi}} (x + 0.044715x^3) \right] \right) \quad (3);$$

[0024] 根据式(1)和式(2),发现GELU是GLU的一种特殊情况,即 f 和 g 是恒等函数,将 σ 看作 ϕ ,由于这种相似性的存在,将GLU视为激活函数用于水印攻击领域,并且GLU本身包含非线性,并不依赖于 σ ,即便移除 σ ,GLU依然不包含非线性,因此对GELU进行简化,提出无非线性GeLU;

[0025] 直接在通道维度上将特征映射分成两部分并相乘,SimpleGate通过逐元素乘法来实现,具体公式如下所示:

$$[0026] \quad SimpleGate(X, Y) = X \odot Y \quad (4);$$

[0027] 其中: X 和 Y 是相同大小的特征图,通过将GELU简化为SimpleGate,水印攻击综合能力显著提升。

[0028] 作为本技术方案的进一步限定,通道注意力首先对输入特征进行全局平均池化操作,随后将全局平均池化后的特征传递给两个线性层和ReLU,用于进行特征压缩和映射,最后将特征映射后的结果输入到Sigmoid函数中,得到通道注意力权重,过将计算得到的通道注意力权重与输入特征进行逐元素相乘,实现对输入特征的加权,使用公式表示为:

$$[0029] \quad CA(X) = X * \sigma(W_2 \max(0, W_1 pool(X))) \quad (5);$$

[0030] 其中: X 表示特征图;

[0031] $pool$ 表示将空间信息居合道通道中的全局平均池化操作;

[0032] σ 是一个非线性激活函数Sigmoid;

[0033] W_1 和 W_2 表示全连接层;

[0034] 倘若将通道注意力机制看成一个函数,则式(5)简化为:

$$[0035] \quad CA(X) = X * G(X) \quad (6);$$

[0036] 将 $G(X)$ 看作 $\sigma(W_2 \max(0, W_1 pool(X)))$,保留通道注意力中较为重要的部分,提出简化通道注意力机制具体公式如下所示:

$$[0037] \quad SCA(X) = X * W pool(X) \quad (7)。$$

[0038] 与现有技术相比,本发明的优点和积极效果是:本系统结合了U-Net结构、无非线性通道注意力机制、跳跃连接、无非线性残差网络、卷积和无非线性GeLU激活函数构造了一个无非线性水印攻击网络,在水印攻击能力和水印攻击不可感知性两方面表现上均优越于现有的水印攻击方法。

附图说明

[0039] 图1为本发明的系统结构图。

[0040] 图2为本发明的无非线性残差块。

[0041] 图3为本发明的无非线性GeLU激活函数。

[0042] 图4为本发明的注意力机制简化对比图。

具体实施方式

[0043] 下面结合附图,对本发明的一个具体实施方式进行详细描述,但应当理解本发明的保护范围并不受具体实施方式的限制。

[0044] 本发明包括以下步骤:

[0045] 采用经典的U-Net架构和跳跃连接,同时使用融入通道注意力的残差网络对RGB图像进行水印攻击任务;U-Net架构不会对网络性能造成阻碍。

[0046] 在非线性水印攻击网络、残差网络和通道注意力中均使用无非线性残差块,进行非线性的去除处理,获得了更高的水印攻击效果;

[0047] 无非线性残差块,接收输入特征图,并通过一系列的卷积操作提取特征,卷积层用于进行卷积操作;转置卷积用于进行上采样操作,非线性水印攻击网络使用了这两个层来进行特征图的下采样和上采样;

[0048] ReLU激活函数在计算机视觉中被广泛使用,然而目前最先行的一些视觉修复方法倾向于使用GeLU将ReLU替换掉,所以本算法中没有采用ReLU作为激活函数,而是采用GeLU(非线性激活函数),并且对GeLU进行非线性的去除,有效解决的训练过程中的梯度消失问题,增强了神经元的感知能力,提高了对输入数据的表达能力;最终,无非线性GeLU的引入使得非线性水印攻击网络在保持水印攻击能力的同时,获得了更高的不可感知性;

[0049] 考虑到计算机视觉中Transformer的近期流行,无非线性残差块内引入注意力机制,并且对通道注意力机制进行了进一步优化,提出一种简化的不含有非线性的通道注意力,该注意力机制在水印攻击工作中可以用于增强模型对输入特征的关注,从而提取更有代表性的特征。

[0050] 无非线性残差块使用门控线性单元进行非线性去除工作;

[0051] GLU(Gated Linear Units,门控线性单元)的提出源于Transformer模型中的一种结构设计,是一种常用的神经网络模块,被广泛应用在自然语言处理、计算机视觉等领域。它结合线性变换和门控函数,可有效地提取序列中的重要特征信息,防止过拟合现象的发生。GLU在一些计算机视觉领域的广泛应用,这意味着GLU也许能够应用于水印攻击领域。可定义为:

$$[0052] \quad \text{GLU}(X, f, g, \sigma) = f(X) \odot \sigma(g(X)) \quad (1);$$

[0053] 其中： X 表示特征映射；

[0054] f 和 g 是线性变换器；

[0055] σ 是一个非线性激活函数

[0056] $\odot$ 表示逐元素乘法。

[0057] 非线性激活函数可表示为：

[0058] $GELU(x) = x\phi(x)$ (2)；

[0059] 其中： ϕ 表示标准正态分布的积累分布函数；

[0060] GELU可以通过以下方式进行近似实现：

[0061] $0.5x \left(1 + \tanh \left[\sqrt{\frac{2}{\pi}} (x + 0.044715x^3) \right] \right)$ (3)；

[0062] 根据式(1)和式(2)，可发现GELU是GLU的一种特殊情况，即 f 和 g 是恒等函数，将 σ 看作 ϕ ，由于这种相似性的存在，可将GLU视为激活函数用于水印攻击领域，并且GLU本身包含非线性，并不依赖于 σ ，即便移除 σ ，GLU依然不包含非线性，因此对GELU进行简化，提出无非线性GeLU；

[0063] 直接在通道维度上将特征映射分成两部分并相乘，于GELU相比，SimpleGate通过逐元素乘法来实现，不会给网络结构带来更多的复杂度，并且保证了网络的水印攻击能力和不可感知性，具体公式如下所示：

[0064] $SimpleGate(X, Y) = X \odot Y$ (4)；

[0065] 其中： X 和 Y 是相同大小的特征图，通过将GELU简化为SimpleGate，水印攻击综合能力显著提升。

[0066] 简化了GELU，使用SimpleGate代替GELU在水印攻击工作中的作用，并且取得了不错的攻击表现。但是在通道注意力模块中的非线性依然存在，接下来我们将讨论它的简化。

[0067] 将CA(通道注意力)运用到水印攻击任务当中有利于引导模型学习含水印图像的关键特征，提高水印攻击的性能。如图4所示，通道注意力首先对输入特征进行全局平均池化操作，随后将全局平均池化后的特征传递给两个线性层和ReLU，用于进行特征压缩和映射，最后将特征映射后的结果输入到Sigmoid函数中，得到通道注意力权重，过将计算得到的通道注意力权重与输入特征进行逐元素相乘，实现对输入特征的加权，使用公式可以表示为：

[0068] $CA(X) = X * \sigma(W_2 \max(0, W_1 pool(X)))$ (5)；

[0069] 其中： X 表示特征图；

[0070] $pool$ 表示将空间信息居合道通道中的全局平均池化操作；

[0071] σ 是一个非线性激活函数Sigmoid；

[0072] W_1 和 W_2 表示全连接层;

[0073] 倘若将通道注意力机制看成一个函数,则式(5)可简化为:

[0074] $CA(X) = X * G(X)$ (6);

[0075] 将 $G(X)$ 看作 $\sigma(W_2 \max(0, W_1 pool(X)))$,保留通道注意力中较为重要的部分,提出简化通道注意力机制具体公式如下所示:

[0076] $SCA(X) = X * Wpool(X)$ (7)。

[0077] 将式(7)运用到了非线性残块中,实现非线性水印攻击系统。

[0078] 深度学习攻击的比较:

[0079] 表1 BG和NNLG与深度学习水印攻击方法的比较

[0080]		FCNDA	Geng	CWAN	Li	本算法
	PSNR	25.3023	26.7405	30.3979	28.3265	37.3380
	BER	0.2120	0.1563	0.1797	0.0976	0.3205
	SSIM	0.9690	0.9821	0.8485	0.9884	0.9642

[0081] 实验结果表明非线性水印攻击系统在不可感知性方面表现出色。PSNR达到了37.3380,均高于其他深度学习水印攻击方法的数值结果。能够做到更加难以察觉的水印攻击。

[0082] 此外,在水印攻击效果方面,本系统的BER则达到了0.3205,远高于其他方法,在水印攻击效果上具有显著的优势。

[0083] 跨数据集攻击效果比较

[0084] 表2 SIDD数据集上的攻击效果比较

[0085]		高斯噪声	JPEG 70	椒盐噪声	均值滤波	本算法
	PSNR	30.0336	34.611	24.9948	32.247	38.5737
	BER	0.0285	0.0473	1541	0.1319	0.1495

[0086] 表3 ImageNet数据集上的攻击效果比较

[0087]		高斯噪声	JPEG 70	椒盐噪声	均值滤波	本算法
	PSNR	30.2335	32.9982	24.7453	26.7842	37.6686
	BER	0.0172	0.0412	0.1322	0.1941	0.2244

[0088] 为了更全面地评估,进行跨数据集的攻击效果比较,使用了SIDD数据集和ImageNet数据集。

[0089] 根据表2的结果,在SIDD数据集上分别使用传统水印攻击方法以及本系统进行了水印攻击任务。本系统仍然优于绝大多数传统水印攻击方法,并且在SIDD数据集上保持了

38.5737的PSNR值,有效保证了攻击后含水印图像的实用价值。表3展示了在ImageNet数据集上的攻击效果实验结果。本系统在攻击效果方面略高于在SIDD数据集上的表现,BER分别达到了0.2244。同时,它们仍然具有比传统攻击方法更高的不可感知性,PSNR达到了37.6686。

[0090] 综上所述,本系统在不同的数据集上均具有高不可感知性和不错的攻击效果。这进一步证明了无非线性水印攻击网络的攻击能力和适用性。

[0091] 本发明的处理流程如下:

[0092] 输入为含水印图像

[0093] 经过卷积操作对输入进行特征提取,

[0094] 无非线性残差块实现了无非线性残差网络和无非线性通道注意力机制的融合,用于学习输入特征的高级表示。

[0095] 下采样操作有助于减少特征图的尺寸,对输入图像进行降维,并提取出更丰富的特征信息。

[0096] 通过无非线性GeLU定义了一个简单的神经网络门的类。这个门的功能是将输入的张量分成两部分,然后将这两部分相乘,最后返回结果。通过构建这样的简单门模型,可以实现对输入数据的拆分和操作,从而使得攻击比较简单地直接对水印进行篡改或破坏。

[0097] 转置卷积的作用是上采样,它将输入特征图的尺寸扩大。在水印攻击任务中,这段转置卷积的作用可能是为了将特征图的尺寸恢复到原始大小,并且减少通道数。这样做的目的是在水印攻击中提取更准确和更精细的特征。

[0098] 随后通过几个卷积操作进行通道缩减。

[0099] 为了减少信息丢失,在其中穿插了两个跳跃连接。

[0100] 最后将神经网络的输出与输入进行结合,即将上采样后的特征图与原始输入图像相加,以保留输入图像的一些细节信息。这有助于改善水印攻击的质量,使得更加清晰和真实。

[0101] 以上公开的仅为本发明的具体实施例,但是,本发明并非局限于此,任何本领域的技术人员能思之的变化都应落入本发明的保护范围。

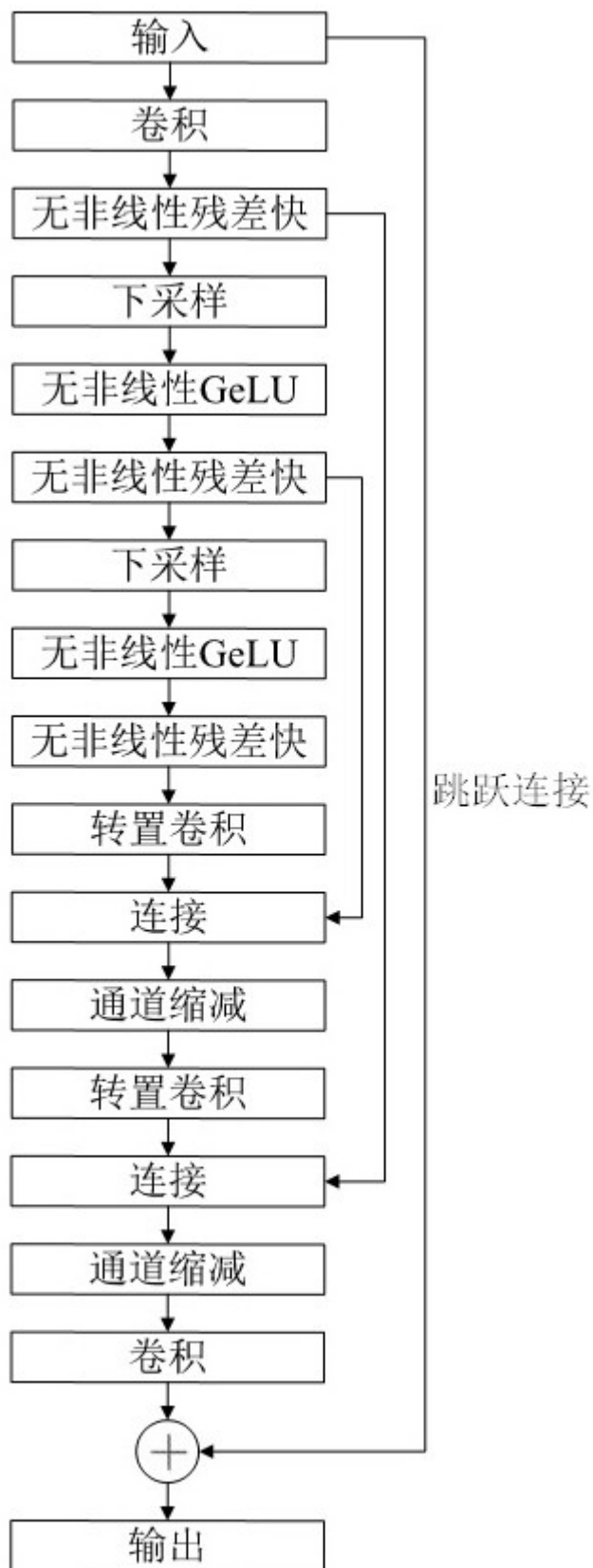


图 1

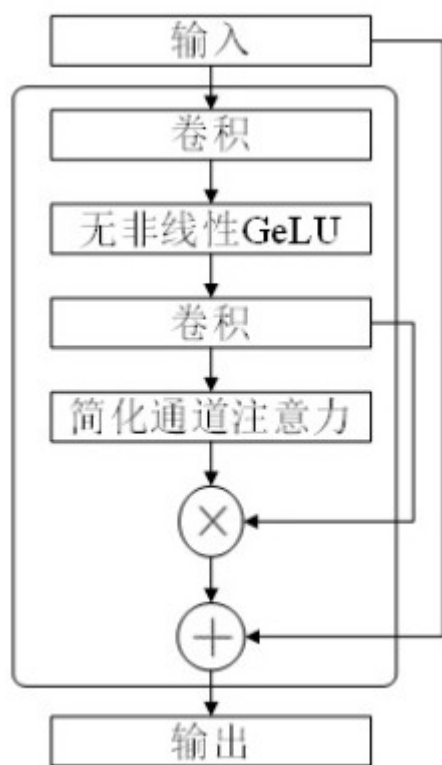


图 2

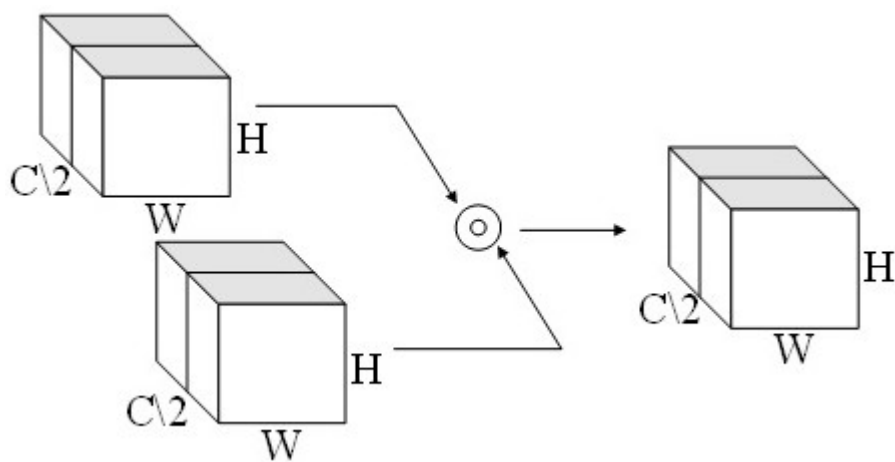


图 3

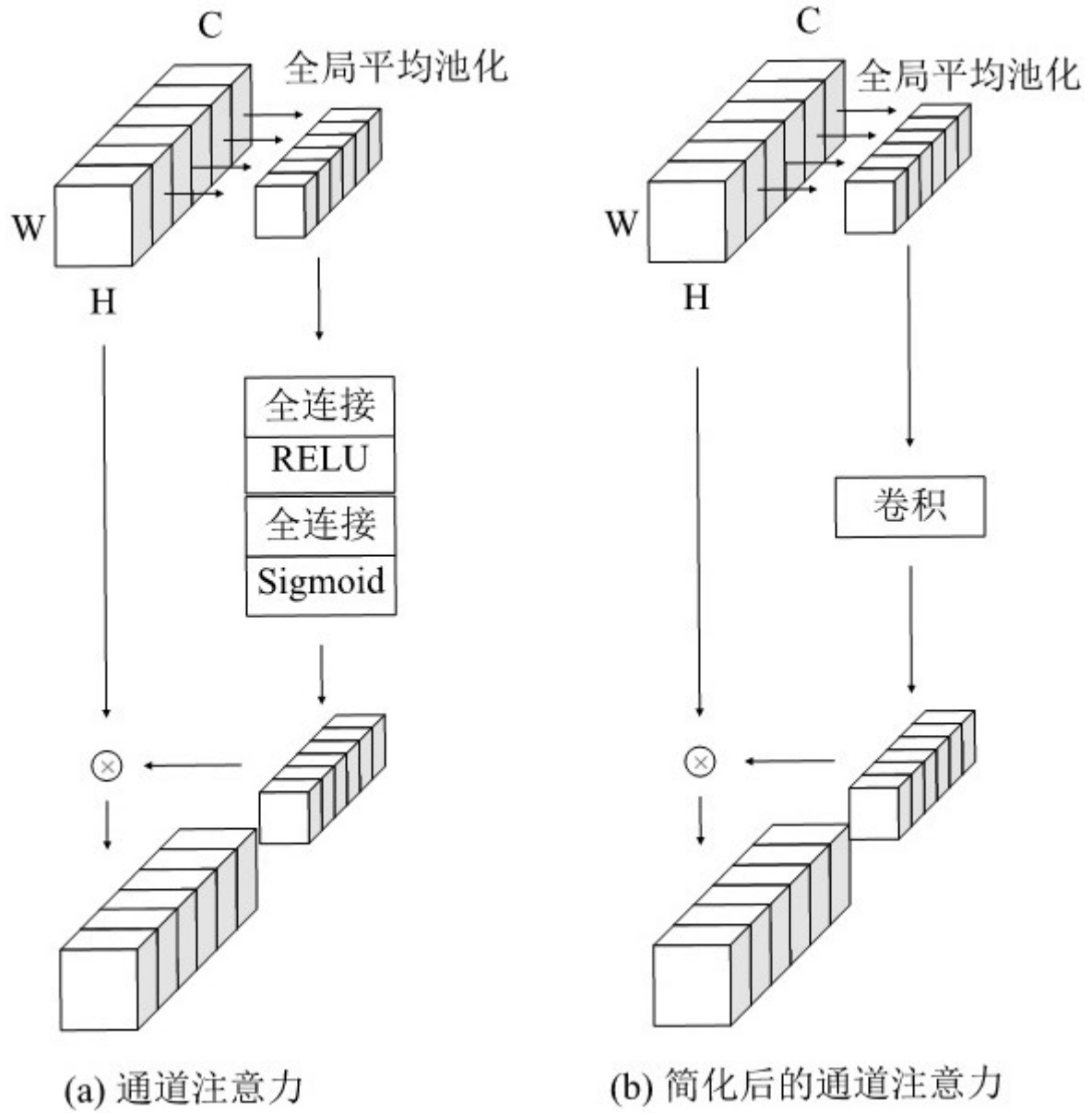


图 4