

# 基于注意力残差网络的卫星图像翻译方法

王晋宇, 张长弓, 杨海涛\*, 冯博迪, 李高源, 高宇歌

航天工程大学航天信息学院, 北京 101416

**摘要** 卫星图像翻译是生成对抗网络的重要应用场景之一, 现有的卫星图像翻译存在生成质量低、泛化能力弱、计算代价高等问题。在循环生成对抗网络的基础上, 通过设计轻量化的注意力残差模块来提高图像翻译质量、减小模型的参数计算量, 同时引入最小二乘损失来提高训练过程的稳定性。实验结果表明, 所提方法在保持高训练稳定性、低模型计算量的前提下, 在卫星图像翻译任务中具有较好的翻译质量。

**关键词** 图像处理; 循环生成对抗网络; 注意力机制; 分离卷积; 最小二乘损失

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202259.0210017

## Satellite Image Translation Method Based on Attention Residual Network

Wang Jinyu, Zhang Changong, Yang Haitao\*, Feng Bodi, Li Gaoyuan, Gao Yuge

School of Space Information, Space Engineering University, Beijing 101416, China

**Abstract** Satellite image translation is one of the important application scenarios of generative adversarial networks. The existing satellite image translation has the problems of low generation quality, weak generalization ability, and high computational cost. Based on the cycle generative adversarial network, a lightweight attention residual module is designed to improve the image translation quality and reduce the parameter computation of the model. At the same time, the least squares loss is introduced to improve the stability of the training process. The experimental results show that the proposed method has good translation quality in satellite image translation tasks while maintaining high training stability and low model computation.

**Key words** image processing; cycle generative adversarial network; attention mechanism; separate convolution; least squares loss

## 1 引言

图像翻译是指不同图像域之间通过建立映射关系而实现的跨域转换。从任务角度来说, 图像翻译可分为语义翻译<sup>[1-3]</sup>和风格翻译<sup>[4-6]</sup>两种。前者注重语义与真实图像间的转换, 如将文字、分割图、地图等语义信息转换为真实图像; 后者注重真实图像间的转换, 如物种转换、昼夜转换、肤色转换等。卫星图像翻译就属于图像翻译的一种, 通过建立多域间映射来实现地图→卫图的语义翻译

及可见光→合成孔径雷达(SAR)、可见光→可见光的风格翻译等。卫星图像翻译具有广泛的应用背景, 如数据扩增、图像分割、城市规划、图像伪造等。

基于生成对抗网络的图像翻译方法是一个热门话题。最初, Gatys等<sup>[7]</sup>将深度学习理论应用于图像翻译任务中, 借助 Gram 矩阵试图将图像的“内容”与“风格”分开, 从而实现风格的转换。Nikulin等<sup>[8]</sup>讨论了文献<sup>[7]</sup>中不同超参数对翻译效果的影响。随后, Johnson等<sup>[9]</sup>在 Gatys等<sup>[7]</sup>的基础上提出

收稿日期: 2021-07-29; 修回日期: 2021-08-28; 录用日期: 2021-09-02

通信作者: \*767153436@qq.com

了一种基于感知损失的翻译方法,该方法使用了残差模块。但上述方法所用到的数据集多为未经匹配的数据样本,生成图像的质量普遍较低。Isola 等<sup>[10]</sup>提出了一种基于条件学习的图像翻译方法,利用配对的数据样本实现了语义图像转换,该方法虽然提高了图像翻译的质量,但对数据集的要求较高,同时训练时间较长,不适用于大规模数据条件下的翻译任务。受对偶学习生成对抗网络<sup>[11]</sup>的启发,Zhu 等<sup>[12]</sup>设计了一种基于循环一致性的图像翻译方法,该方法可以从无需匹配的数据集中实现风格转换,但存在生成细节少、泛化能力差等方面的问题。

在卫星图像翻译领域,基于生成对抗网络的图像翻译已经有了一定的应用,如文献<sup>[13-15]</sup>通过条件式生成对抗网络进行卫星图像翻译,文献<sup>[16-17]</sup>在基于循环一致性基础上进行了改进,成功实现了可见光到 SAR 图像的风格翻译。地图语义翻译方面,CycleGAN、Pixpix 等方法实现了电子地图至卫星地图之间的转换。文献<sup>[18]</sup>利用 Deepfake 技术,在以华盛顿地图为底图的基础上可以生成北京和西雅图风格的伪造卫星图像,但该方法所利用的卫星图像数据分辨率较低。同时上述方法通常只针对单一应用场景,模型的泛化能力较差。

利用注意力机制对生成网络的生成效果进行改进在国内外研究中有许多应用。具有代表性的 SAGAN 网络<sup>[19]</sup>第一次将注意力引入 GAN 网络中,并将 inception score (IS) 指数提高到了 52.52。之后 BigGAN 网络<sup>[20]</sup>在 SAGAN 网络的基础上进一步将 IS 提高到了 98.8。文献<sup>[21]</sup>利用自注意力机制设计了一种渐进式的图像合成方法,DRB-GAN 网络<sup>[22]</sup>通过注意力风格码的生成实现了图像艺术风格的迁移。王越等<sup>[23]</sup>利用

金字塔注意力机制和 GAN 网络的结合进行水下图像增强,文献<sup>[24]</sup>利用注意力机制进行行人轨迹预测等。

综上所述,本文提出了一种基于循环一致性的无监督遥感图像翻译方法 SA-CycleGAN。在 CycleGAN 网络的基础上,设计了一种分离注意力残差模块来提高图像的生成质量,同时利用最小二乘损失增强训练过程的稳定性。作为一种无监督的方法,所提方法无需一一匹配的数据集,且通过轻量化的设计,所提方法在拥有更好生成质量的同时较大程度减轻了计算代价。

## 2 生成对抗网络

GAN 网络是一种基于对抗思想的网络架构,如图 1 所示,主要部分包含生成器 G 和鉴别器 D。D 用于判断输入图像是否为真,对于真实图像,D 输出高分,反之输出低分;G 用于生成图像并输入 D 判别。随着训练参数的更新,G 不断学习如何能够“欺骗”D,从而使生成样本越来越接近真实图像。为了方便表达,用 D、G 分别表示输入样本经过生成器 G 和鉴别器 D 处理后的结果。GAN 网络的目标函数为

$$\min_G \max_D V(D, G) = E_{x \sim P_{\text{data}}(x)} \log D(x) + E_{z \sim P_{z(z)}} \log \{1 - D[G(z)]\}, \quad (1)$$

式中: $E_{x \sim P_{\text{data}}(x)} \log D(x)$ 是真实数据在 D 中的期望, $x \sim P_{\text{data}}(x)$ 表示采样于真实数据的输入样本; $E_{z \sim P_{z(z)}} \log \{1 - D[G(z)]\}$ 是生成数据在 D 中的期望, $z \sim P_{z(z)}$ 表示采样于噪声数据的输入样本; $D[\cdot]$ 代表经过 D 后的输出分数。训练过程中的对抗思想体现在  $E_{z \sim P_{z(z)}} \log \{1 - D[G(z)]\}$  部分。GAN 在实际中采用先固定 D 训练 G,后固定 G 训练 D 的方法。因此,当固定 D 时,目标在于最大化  $V(D, G)$ ;固定 G 时,目标在于最小化  $V(D, G)$ 。

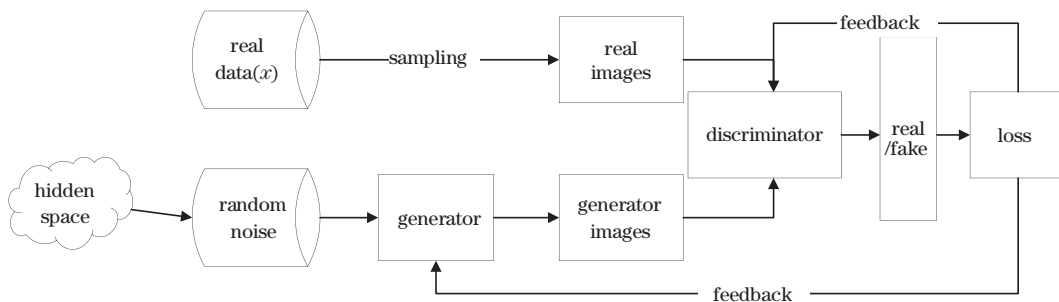


图 1 生成对抗网络架构

Fig. 1 Generating adversarial network architecture

### 3 所提方法

#### 3.1 分离注意力残差模块

图 2 为所设计的用于提升生成模型特征提取能力的分离注意力残差模块(SA-Restnet)。残差模块是 CycleGAN 网络中的核心部分,跳跃连接方式增加了数据的传输支路,从而增强了信息流的传递速度,也使信息在由浅入深的传播过程中不易发生梯

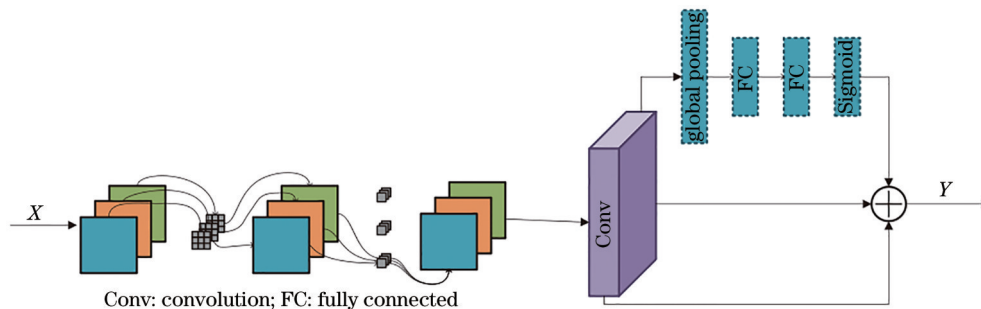


图 2 分离注意力残差模块

Fig. 2 Separating attention residual module

因此,首先利用了 SENet 通道注意力机制对残差模块进行设计。SENet 可以有效提取网络模型各通道间的权重信息,通过输出一组权重向量来增强有用通道并抑制冗余通道。SENet 由 Squeeze 和 Excitation 两部分组成,其中 Squeeze 可以将  $C \times H \times W$  的特征图通过全局池化的方法压缩为  $C \times 1 \times 1$ 。接着 Excitation 利用两个全连接层进行通道特征学习,并通过 Sigmoid 激活函数输出权重向量。

对于残差模块来说,希望对输入数据的低维特征首先进行初步提取,从而提高之后残差学习的效率。CycleGAN 网络使用卷积核大小为 3 的二维卷积单元来达到这一目的,但这种方法属于一种粗提取,会给生成器带来大量的计算参数,性价比不高。分离卷积是一种通过减小参数量来提高计算速度的有效方法<sup>[25]</sup>,首先利用深度分组卷积对空间维度进行信息提取,之后利用  $1 \times 1$  逐点卷积对通道维度进行信息提取,从而建立通道与空间联合相关的映射关系,这种方法在保证信息提取有效性的前提下可以大幅降低模型参数量,从而提高模型的运算速度。

#### 3.2 网络架构

所提方法使用基于循环一致性的样本生成策略,实现两种图像域之间的转换,如图 3 所示,其中 G 可以使图像域 A 中的样本转换到图像域  $\tilde{A}$ ,F 可以使图像域 B 中的样本转换到图像域  $\tilde{B}$ 。在这个过

度弥散和网络退化问题。残差网络解决了训练过程中的过拟合及崩溃问题,同时提升了模型训练的稳定性,但并未有效提升图像的生成质量,原因在于简单地进行了残差网络堆积带来了大量的冗余通道,如 CycleGAN 网络在  $256 \times 256$  的图像翻译任务中使用的 9 层级联残差模块共有 2304 个特征通道参数需要学习。过多的冗余通道不利于获取更为丰富的信息特征,从而制约了图像生成质量。

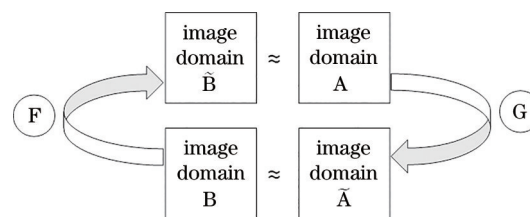


图 3 循环一致原理

Fig. 3 Consistent principle of circulation

程中,循环训练的目的是使  $\tilde{A} \approx B, \tilde{B} \approx A$ ,从而达到跨域翻译的目的。值得注意的是,在无监督条件下,域 A 和域 B 间的数据无需进行匹配,这样大大减小了数据集制备的难度。

根据循环一致性原理得到的网络架构如图 4 所示。不同于传统生成器的高斯分布输入,该架构中 G 和 F 分别输入卫星域 A 和地图域 B 的图像数据,G 负责建立  $A \rightarrow B$  的映射,F 负责建立  $B \rightarrow A$  的映射, $D_A$  和  $D_B$  为两个鉴别器。首先,G、 $D_B$  和 F、 $D_A$  构成两个传统 GAN 网络,其中  $D_A$  用来区分真实图像 RealA 和生成图像 FakeA, $D_B$  用来区分真实图像 RealB 和生成图像 FakeB。其次,为了学习 A 到 B 之间的循环映射,整个网络存在  $RealA \rightarrow FakeB \rightarrow RecA$  和  $RealB \rightarrow FakeA \rightarrow RecB$  两条重建网络,RecA 和 RecB 为经过循环生成过程后的重建图像。重建网络的中间层 FakeA 和 FakeB 即为转换器 G 和 F 的生成图像。

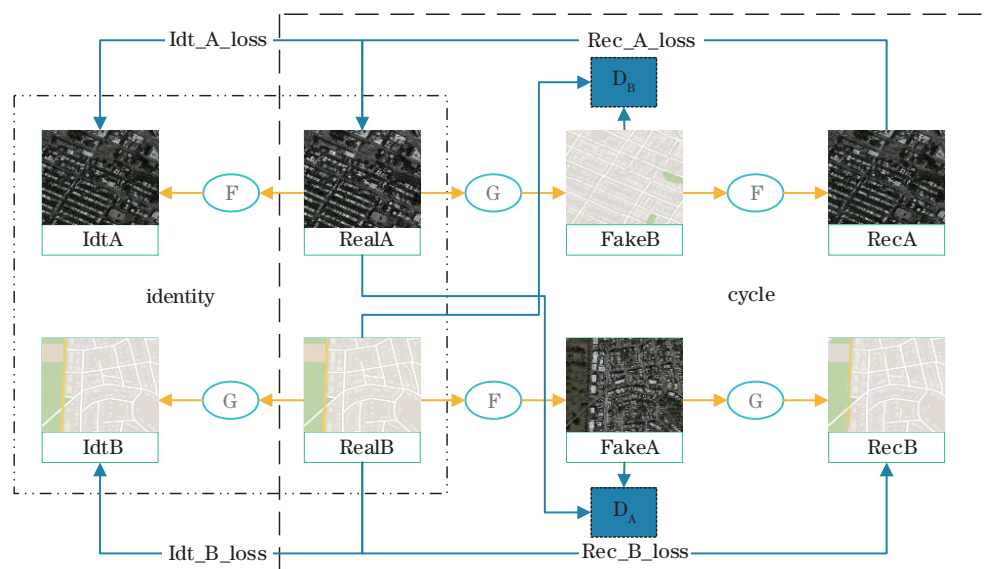


图 4 网络架构

Fig. 4 Network architecture

同时,使用了特征鉴别网络。以 G 为例,G 本身可以实现域 A 到域 B 的转换,而输入为域 B 的图像时,输出也应当为域 B。利用这一原理,对 G 和 F 使用特征鉴别,从而提高生成图像的质量。IdtA 和 IdtB 分别为 F 和 G 的鉴别生成图像。

### 3.3 生成器设计

所设计的生成器由编码器、转换器、解码器组成,如图 5 所示,其中转换器使用 9 层级联的 SA-Restnet 模块代替 CycleGAN 中的残差模块;编码器由 3 层卷积网络构成,负责将  $3 \times 256 \times 256$  的原始输入图像编码为  $256 \times 64 \times 64$  的特征图;转换器由 9 层 SA-Restnet 组成,负责学习图像域转换任务中的主

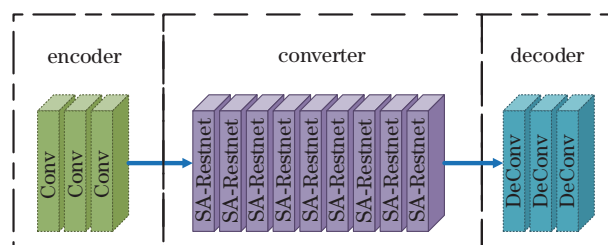


图 5 生成器

Fig. 5 Generator

要参数,输出尺寸为  $256 \times 64 \times 64$ ;解码器将转换器的输出结果通过 3 层反卷积变换为输出图像,输出尺寸为  $3 \times 256 \times 256$ 。整个生成器的网络配置如表 1 所示。

表 1 生成器网络配置

Table 1 Generator network configuration

Type	Procedure	Name	Number	Input size	Output size	Normalization	Activation function	Padding
Encoder	1	Conv	1	$3 \times 256 \times 256$	$64 \times 256 \times 256$	In	ReLU	Yes
	2	Conv	1	$64 \times 256 \times 256$	$128 \times 128 \times 128$	In	ReLU	No
	3	Conv	1	$128 \times 128 \times 128$	$256 \times 64 \times 64$	In	ReLU	No
Converter	4	SeparConv	9	$256 \times 64 \times 64$	$256 \times 64 \times 64$	In	ReLU	Yes
	5	SERestnet	9	$256 \times 64 \times 64$	$256 \times 64 \times 64$	In	ReLU	No
Decoder	6	DeConv	1	$256 \times 64 \times 64$	$128 \times 128 \times 128$	In	ReLU	No
	7	DeConv	1	$128 \times 128 \times 128$	$64 \times 256 \times 256$	In	ReLU	No
	8	DeConv	1	$64 \times 256 \times 256$	$3 \times 256 \times 256$	In	ReLU	Yes

### 3.4 损失函数设计

首先,对于 G、 $D_B$  和 F、 $D_B$  两个对抗结构,需要构建两个对抗损失。原始 CycleGAN 网络采用交叉熵损失来衡量生成样本和真实样本之间的距离,然而

这种方法下的训练过程稳定性较差,损失的跳跃性强,容易出现模型崩溃的问题。因此,将原始交叉熵损失改进为最小二乘损失,最小二乘损失可以使判定为真且远离决策边界的假样本继续得到优

化,从而有效提升了图像生成质量及训练的稳定性。最小二乘损失函数下的对抗损失为

$$L_{adv}(F, D_A, \text{RealA}, \text{RealB}) = E_{a \sim P_{\text{RealA}}} \log[D_A(\mathbf{a})] + E_{b \sim P_{\text{RealB}}} \log\{1 - D_A[F(\mathbf{b})]\}, \quad (2)$$

式中: $F$ 为经过转换器 $F$ 后的结果; $D_A$ 为经过鉴别器 $D_A$ 后的结果; $\mathbf{a}$ 、 $\mathbf{b}$ 分别表示采样于图像域A和图像域B的张量。

$$L_{adv}(G, D_B, \text{RealB}, \text{RealA}) = E_{b \sim P_{\text{RealB}}} \log[D_B(\mathbf{b})] + E_{a \sim P_{\text{RealA}}} \log\{1 - D_B[G(\mathbf{a})]\}, \quad (3)$$

式中: $G$ 为经过转换器 $G$ 后的结果; $D_B$ 为经过鉴别器 $D_B$ 后的结果。

对于两条重建网络来说,利用L1范数作为距离计算方法构成的两个重建损失为

$$L_{\text{RecA}}(G, F, \text{RealA}) = E_{a \sim P_{\text{RealA}}} \left\{ \|F[G(\mathbf{a})] - \mathbf{a}\|_1 \right\}, \quad (4)$$

$$L_{\text{RecB}}(F, G, \text{RealB}) = E_{b \sim P_{\text{RealB}}} \left\{ \|G[F(\mathbf{b})] - \mathbf{b}\|_1 \right\}. \quad (5)$$

根据重建损失及循环一致原理,建立的循环损失为

$$L_{\text{cyc}}(G, F) = L_{\text{RecA}}(G, F, \text{RealA}) + L_{\text{RecB}}(F, G, \text{RealB}). \quad (6)$$

循环损失从数值上为两个重建损失之和。同时,在L1范数条件下的 $G$ 和 $F$ 的特征损失函数为

$$L_{\text{Idt}}(G, F) = L_{\text{IdtA}}(F, \text{RealA}) + L_{\text{IdtB}}(G, \text{RealB}) = E_{a \sim P_{\text{RealA}}} \left[ \|F(\mathbf{a}) - \mathbf{a}\|_1 \right] + E_{b \sim P_{\text{RealB}}} \left[ \|F(\mathbf{b}) - \mathbf{b}\|_1 \right]. \quad (7)$$

综上所述,对于损失函数来说,总共使用到了两个对抗损失、一个循环损失和一个特征损失。因此所提方法的目标函数为

$$L(G, F, D_A, D_B) = L_{\text{RecA}}(G, F, \text{RealA}) + L_{\text{RecB}}(F, G, \text{RealB}) + lL_{\text{cyc}}(G, F) + mL_{\text{Idt}}(G, F), \quad (8)$$

式中: $l$ 和 $m$ 为经验参数。

## 4 实验与结果分析

### 4.1 数据集制备

以地图语义翻译为例,验证所提SA-CycleGAN网络的翻译效果,并在风格翻译任务中验证泛化能力。因此,共建立了4个数据集,如表2所示。

数据集1属于语义翻译数据集,用于验证地图语义翻译效果,采用了文献[12]中的数据源进行语义地图与卫星图像之间的翻译实验。其余3个为风格翻译数据集,用于验证泛化能力。数据集2使用了RSI-CB数据集,用于实现厂房屋顶颜色的转换。数据集3和4使用了RSI-CB数据集和GF3数据,用于实现SAR图像和可见光图像之间的风格转换,其中SAR图像部分由两景高分三号数据裁剪而成。所用到的SAR图像缩略图如图6所示。

表2 数据集内容

Table 2 Content of dataset

Type of translation	Dataset	Terrestrial	DataA	DataB	Data source	Size	Data volume(A)	Data volume(B)
Semantic Translation	1	City	Maps	Images	CycleGAN <sup>[12]</sup>	256	1098	1096
Style Translation	2	Factory	Images	Images	RSI-CB	256	2511	2445
	3	Vegetation	Images	SAR	RSI-CB、GF-3	256	2358	848
	4	Farmland	Images	SAR	RSI-CB、GF-3	256	2491	656

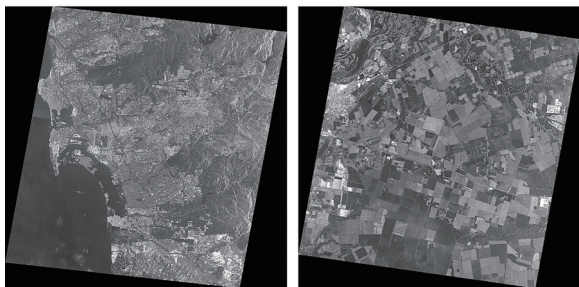


图6 SAR图像数据

Fig. 6 Image data of SAR

所用数据经筛选、切片和裁切处理后的可视化数据集如图7所示。

### 4.2 实验过程

使用语义翻译数据集1进行地图翻译任务,同时将所提方法与主流方法Pixpix、CycleGAN进行对比试验。数据集1具有较为丰富的色彩和纹理特征,其中域A共有1098张地图,域B共有1096张卫星图片。将图片统一到 $256 \times 256$ 尺寸后,通过SA-CycleGAN网络进行训练,同时准备了50对已经配对的图片数据作为验证集。

本实验所用的环境为 Windows 10 系统、AMD Ryzen7 5800X 处理器、32GB RAM、NVIDIA GeForce RTX3090 显卡。所用到的深度学习框架为 Pytorch1.8, CUDA 版本为 11.1。对于训练过程, 每组方法共训练 80 轮, 优化器选用 Adam (beta 为 0.5), 学习率为 0.002, batch size 为 4, 其中 CycleGAN 和所提方法从 50 轮之后采用学习率衰减的方式进行。

### 4.3 对比分析

#### 4.3.1 定性分析

在经过 80 轮训练后, 各种方法生成的效果图如图 8 所示。从地图转卫图角度来说, CycleGAN 和 Pixpix 的失真程度较高。直观上相对于 CycleGAN, 所提方法可以很好地生成水系, 同时相对于 Pixpix, 可以生成更为逼真的道路及建筑。从卫图转地图角度来说, Pixpix 的失真程度较高, CycleGAN 和所提方法对道路的提取能力较好, 但

是 CycleGAN 无法有效提取水系的语义地图信息。因此, 从目视判读的定性角度, 所提方法具有较好的生成效果。

SA-CycleGAN 的无监督学习策略相对于有监督的训练方法来说, 依然具有一定的局限性。例如随着训练轮次的进行, 无监督的图像翻译质量会低于一一匹配的有监督方法。这是因为有监督模型同时学习到了更多的类似于 CGAN 的先验条件信息, 最终使得翻译图像细节的提高, 但这类算法的附加条件会带来更大的计算成本。本实验的 80 轮训练条件下, Pixpix 等有监督学习策略在前期通常能够学习到更多的类别信息, 如水体、林地、道路的色调信息, 但翻译质量不能满足人眼主观认知。而 CycleGAN 等方法由于缺乏监督, 前期可能会学习到错误的类别信息, 从而陷入错误的局部最优解。因此, CycleGAN 网络的提出者也承认成对和非成对训练之间不可消除的差异性, 同时可为模型添加

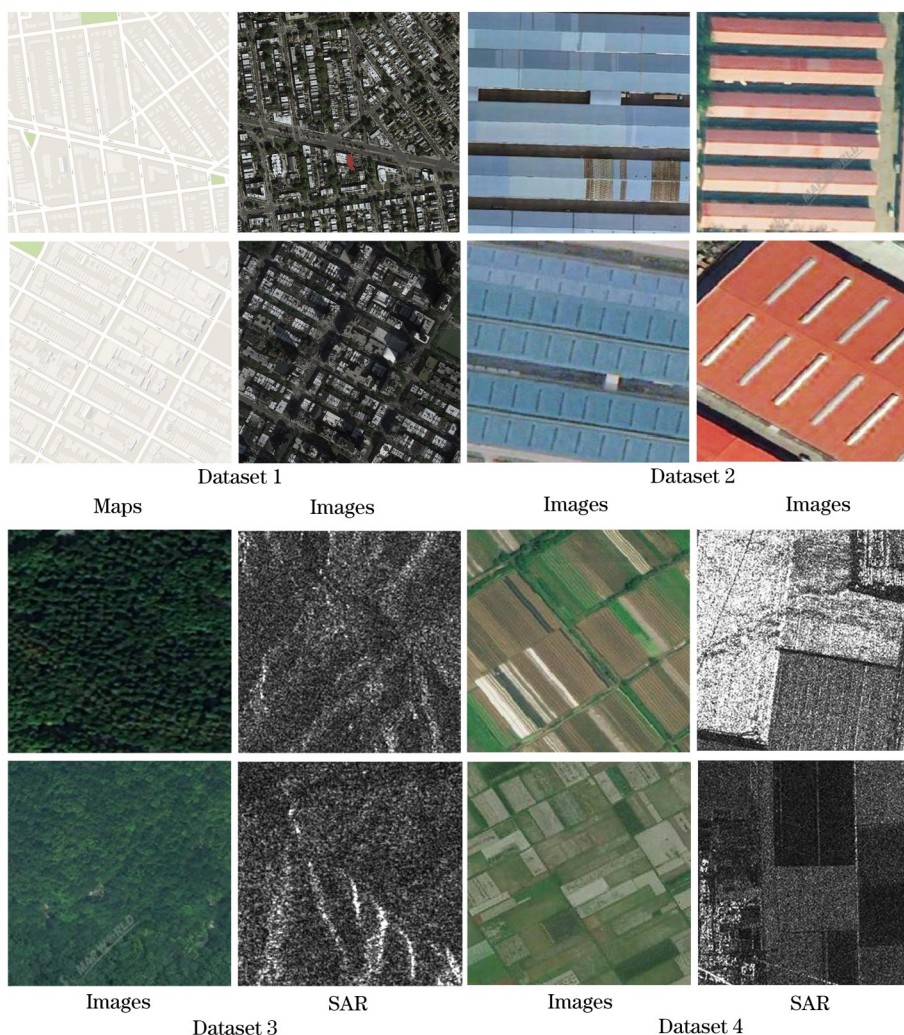


图 7 数据集可视化

Fig. 7 Dataset visualization

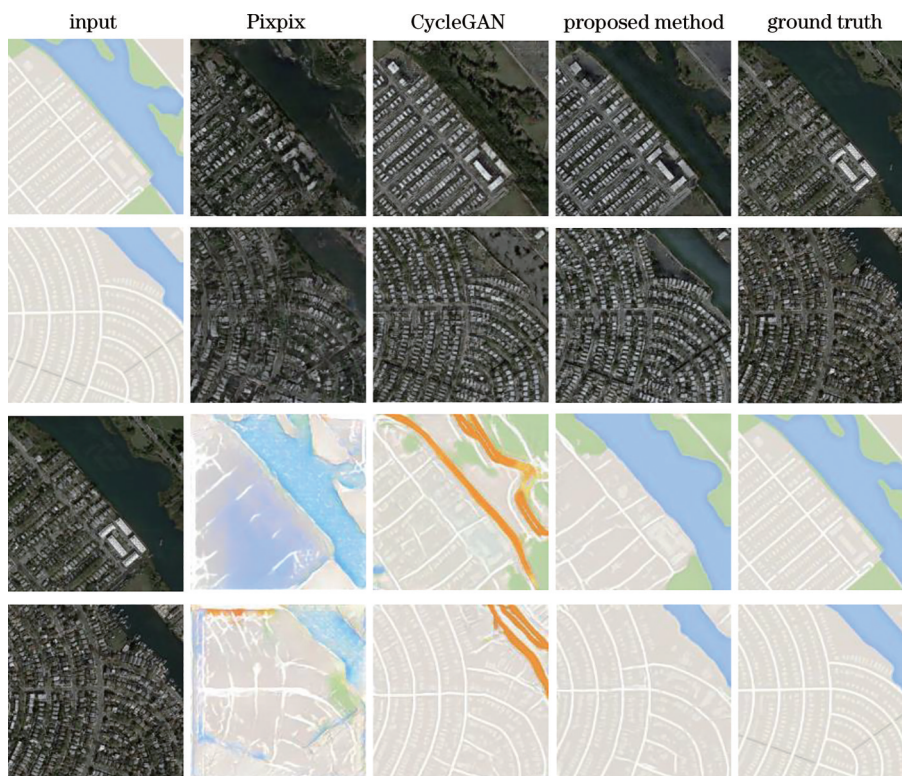


图 8 语义翻译结果

Fig. 8 Results of semantic translation

弱语义监督<sup>[12]</sup>。但在现有条件下,非成对的无监督训练策略更能顾及数据集制备难度和生成效果之间的平衡。

### 4.3.2 定量分析

#### 4.3.2.1 评价指标

所使用的评价指标共有 3 种,分别是峰值信噪比 (PSNR)、结构相似度 (SSIM) 和 Frechet inception distance score (FID)。PSNR 是常用的评价数字图像画质的客观度量方法,用于计算参考图像与真实图像的失真程度。SSIM 是一种模拟人类视觉特征的图像质量评价方法,通过对结构信息的提取来获得图像间的失真程度及相似性。PSNR 和 SSIM 的表达式为

$$R_{\text{PSNR}} = 10 \times \lg \frac{I_{\text{max}}^2}{E_{\text{MSE}}}, \quad (9)$$

式中:  $I_{\text{max}}$  表示遥感图像像元灰度最大值;  $E_{\text{MSE}}$  表示原图像与生成图像灰度值的均方误差。

$$M_{\text{SSIM}}(x, y) = \frac{(2m_x m_y + c_1)(2s_{xy} + c_2)}{(m_x^2 + m_y^2 + c_1)(s_x^2 + s_y^2 + c_2)}, \quad (10)$$

式中:  $m_x, m_y$  为原始样本  $x$  与生成样本  $y$  的灰度值均

值;  $s_x, s_y$  为原始样本  $x$  与生成样本  $y$  的灰度值方差;  $s_{xy}$  为原始样本  $x$  与生成样本  $y$  之间的灰度协方差;  $c_1, c_2$  为常数。

FID 通过 Inception v3 网络进行特征的提取,构成了真实样本和生成样本两个概率分布,通过评价这两个分布之间的距离来达到模型评价的目的。由于优良的噪声抵抗能力, FID 在模型多样性评价方面具有更好的效果。FID 的缺点在于依然没有解决大规模数据集上无法进行模型过拟合评价的问题。

#### 4.3.2.2 生成过程评估

图 9 为 3 种方法在不同训练阶段的 FID 指数变化曲线。从图中可以看出: 所提方法可以保持较低水平的 FID 值; 同时所提方法 FID 下降速度更快, 说明 SA-CycleGAN 可以更快学习到与原图相似的细节特征, 从而提高图像生成质量。图 10(a)、(b) 分别表示 CycleGAN 和所提方法损失函数变化曲线。从图中可以看出: 由于 SA-CycleGAN 模型结构的优势及最小二乘损失的使用, 所提方法可以更快收敛; 同时在生成过程中损失函数的稳定性有了明显的提升。

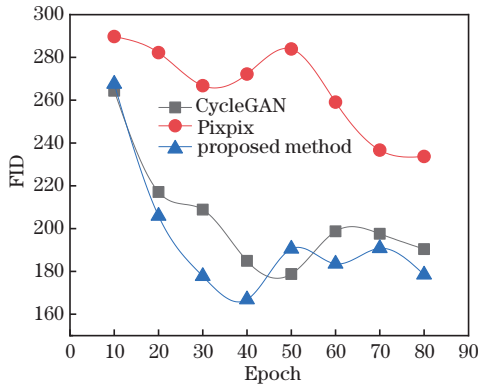


图 9 生成阶段 FID 变化

Fig. 9 FID change during generation phase

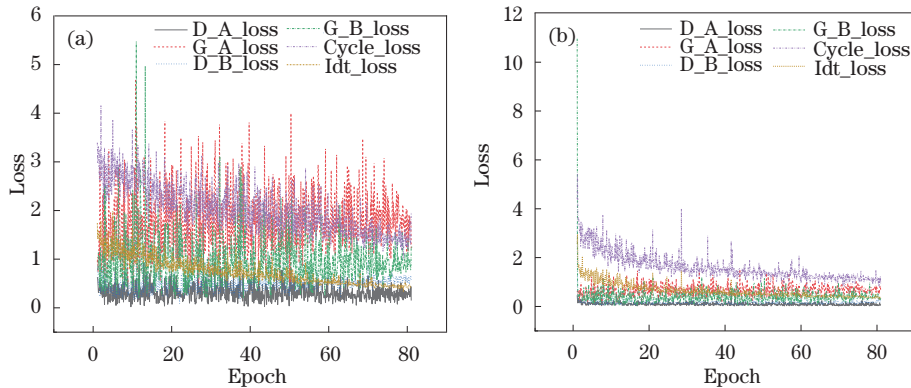


图 10 损失函数曲线。(a) CycleGAN; (b) 所提方法

Fig. 10 Loss function curves. (a) CycleGAN; (b) proposed method

表 3 指标对比

Table 3 Comparison of indicators

method	Parameter / MB	Model size / MB	PSNR		SSIM		FID	
			Map→ Image image	Image→ map	Map→ imag	Image→ map	Map→ imag	Image→ map
CycleGAN	13.7	43.4	14.02	24.98	0.18	0.6	190.36	257.59
Pixpix	54.4	207	<b>14.96</b>	17.09	0.21	0.51	233.69	322.38
Proposed method	<b>4.8</b>	<b>25.8</b>	14.67	<b>26.68</b>	<b>0.22</b>	<b>0.67</b>	<b>178.44</b>	<b>197.56</b>

4.3.3 消融实验

所提方法 SA-CycleGAN 是在原 CycleGAN 基础上,加入了残差注意力模块、分离卷积和最小二乘损失改进而来的。为了验证这 3 种改进措施对网络模型及最终精度的贡献,设计了一组消融实验。

由表 4 客观指标可以看出:残差注意力模块、分离卷积和最小二乘损失 3 者的改进均有助于减小 FID 精度,即提高了图像生成质量。其中分离卷积在有效减小参数量及模型大小的同时,可以带来较高的质量提升,这是由于减小模型参数有助于减轻过拟合,从而提升精度。其他两种改进措施在保持

4.3.2.3 翻译结果评估

表 3 为不同方法的 PSNR、SSIM、FID 指标,其中加粗的为表现最好的指标。从表中可以看出:在地图转卫图任务中 Pixpix 的 PSNR 最优,但 PSNR 仅基于像素值的均方误差进行指标计算,容易造成无法与人眼主观视觉一致的问题;在其余 5 个量化指标上,所提方法都能够达到最优,从多角度证明了鲁棒性。同时,注意到了计算消耗的问题,由于分离卷积的使用,SA-CycleGAN 大幅度降低了参数量及模型大小,与 CycleGAN 相比,模型参数量减小了 65%。这有效提高了运算效率,在更小的计算代价下可以获得更高的图像翻译效果。

实验采用控制变量法进行,将 3 种改进措施依次命名为 A、B、C,共设计了 5 组实验进行对比。实验结果如表 4 所示,其中 ↑ 和 ↓ 分别表示相对于 CycleGAN 的指标数据增减情况,同时 5 组实验所生成可视化图像如图 11 所示。

参数量基本平衡的前提下,也可有效降低 FID 指标。从图 11 主观分析可以看出:3 种改进措施对最终翻译图像质量的贡献也有所差异,如残差注意力模块的加入可以使得模型在卫图→地图的过程中获得较好的道路提取效果,但是对于水系的提取能力较差;分离卷积的加入提高了水系的提取能力,



表 4 消融实验客观评价  
Table 4 Objective evaluation of ablation experiments

Algorithm	Parameter /MB	Model size /MB	FID(map→image)	FID(image→map)
CycleGAN	13.7	43.4	190.36	257.59
CycleGAN+A	13.9 ↑	43.7 ↑	181.68 ↓	208.89 ↓
CycleGAN+B	4.8 ↓	25.5 ↓	180.96 ↓	200.61 ↓
CycleGAN+C	13.7	43.4	181.74 ↓	199.74 ↓
SA-CycleGAN	4.8 ↓	25.8 ↓	178.44 ↓	197.56 ↓

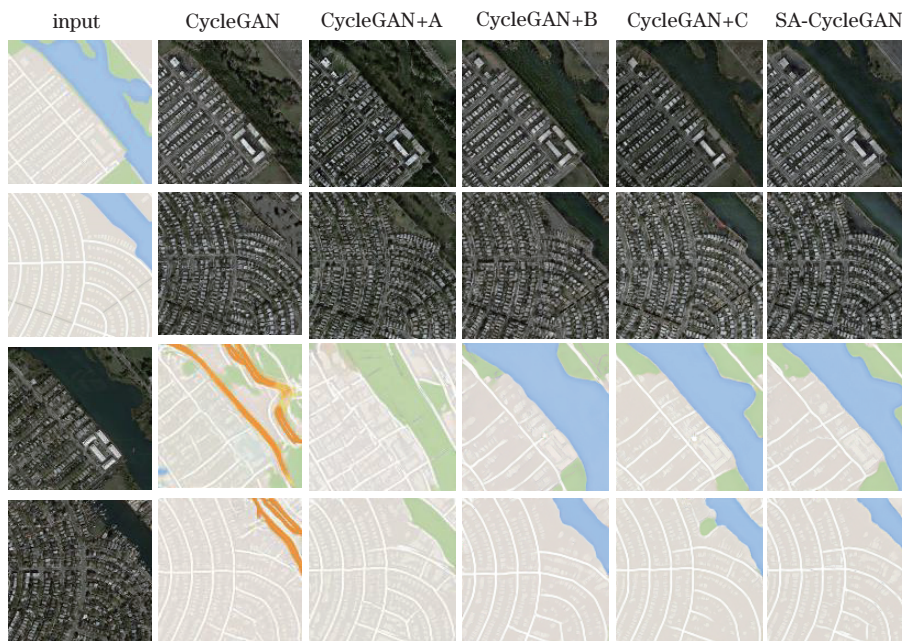


图 11 消融实验主观评价

Fig. 11 Subjective evaluation of ablation experiments

但是无法准确提取边界位置;最小二乘损失的加入在提高训练稳定性的同时进一步提高了信息提取能力,能够较为准确地进行水系边界生成,但翻译质量低于 SA-CycleGAN。因此,从主客观结合的分析过程中得知,所提 SA-CycleGAN 是一种可以提取丰富卫星图像信息的方法。

#### 4.4 模型泛化能力

为了验证所提方法在其他生成场景下的泛化能

力,使用了数据集 2,3 和 4 进行了风格转换,如图 12 所示,其中数据 2 进行基于可见光的厂房风格转换,数据 3 和 4 进行基于可见光和 SAR 之间的风格转换。从图中可以看出,经过风格转换后的图像依然可以保持很好的协调性和清晰度,这为少样本条件下的 SAR、可见光数据量增强等任务提供了一种很好的思路。

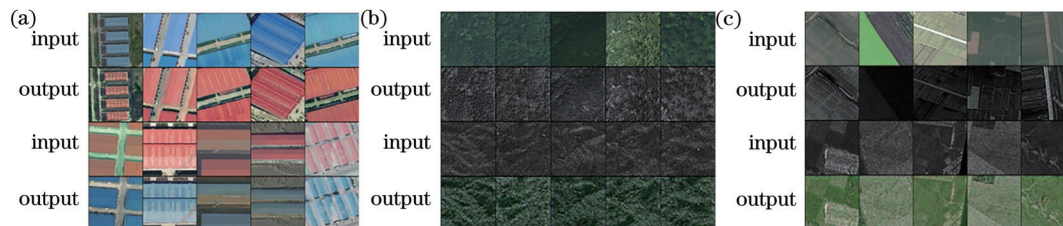


图 12 其他数据集翻译结果。(a)数据集 2;(b)数据集 3;(c)数据集 4

Fig. 12 Translation results for other datasets. (a) Dataset 2; (b) dataset 3; (c) dataset 4

## 5 结 论

提出了一种基于循环一致原理的卫星图像翻译方法 SA-CycleGAN。为了提高生成模型的信息提取能力,设计了一种分离卷积注意力残差模块,该模块通过引入 SENet 注意力机制来抑制残差网络中的冗余通道,从而可以提取更为丰富的图像特征,利用分离卷积模块减小了模型的参数和计算量,同时最小二乘损失的改进有效提高了训练的稳定性。在地图语义翻译任务中对所提方法进行了验证,定性和定量实验结果表明,SA-CycleGAN 在生成过程、生成结果方面相较于 Pixpix 和 CycleGAN 均较好,其中 PSNR、SSIM 和 FID 指标均有显著提升,同时模型参数量相较于 CycleGAN 减少了 65%。最后在可见光→可见光和可见光→SAR 的风格转换任务中验证了所提方法的泛化能力。在今后的研究中,将围绕感兴趣区域的图像翻译任务进行相关研究和实验。

### 参 考 文 献

- [1] Lee J, Kim E, Lee Y, et al. Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 5800-5809.
- [2] Karras T, Laine S, Aila T M. A style-based generator architecture for generative adversarial networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 19263406.
- [3] Zhu B, Ngo C W. CookGAN: causality based text-to-image synthesis[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 5518-5526.
- [4] Shen Y, Yang Q, Yuan YB, et al. Image style migration for target edge clarification[J]. *Advances in Laser and Optoelectronics*, 2021, 58(12):241-253.  
沈瑜,杨倩,苑玉彬,等.目标边缘清晰化的图像风格迁移[J].*激光与光电子学进展*, 2021, 58(12):241-253.
- [5] Anokhin I, Solovov P, Korzhenkov D, et al. High-resolution daytime translation without domain labels [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 7485-7494.
- [6] Li X Y, Zhang S C, Hu J, et al. Image-to-image translation via hierarchical style disentanglement[EB/OL]. (2021-03-02) [2021-08-04]. <https://arxiv.org/abs/2103.01456>.
- [7] Gatys L A, Bethge M, Hertzmann A, et al. Preserving color in neural artistic style transfer[EB/OL]. (2016-06-19) [2021-08-04]. <https://arxiv.org/abs/1602.07188>.
- [8] Nikulin Y, Novak R. Exploring the neural algorithm of artistic style[EB/OL]. (2016-03-13) [2021-08-04]. <https://arxiv.org/abs/1602.07188>.
- [9] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9906: 694-711.
- [10] Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5967-5976.
- [11] Yi Z L, Zhang H, Tan P, et al. DualGAN: unsupervised dual learning for image-to-image translation [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2868-2876.
- [12] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2242-2251.
- [13] Enomoto K, Sakurada K, Wang W M, et al. Image translation between SAR and optical imagery with generative adversarial nets[C]//IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, July 22-27, 2018, Valencia, Spain. New York: IEEE Press, 2018: 1752-1755.
- [14] Fuentes Reyes M, Auer S, Merkle N, et al. SAR-to-optical image translation based on conditional generative adversarial networks: optimization, opportunities and limits[J]. *Remote Sensing*, 2019, 11(17): 2067.
- [15] Niu X, Yang D, Yang K, et al. Image translation between high-resolution optical and synthetic

- aperture radar (SAR) data[J]. International Journal of Remote Sensing, 2021, 42(12): 4758-4784.
- [16] Wang L, Xu X, Yu Y, et al. SAR-to-optical image translation using supervised cycle-consistent adversarial networks[J]. IEEE Access, 2019, 7: 129136-129149.
- [17] Fu S L, Xu F, Jin Y Q. Reciprocal translation between SAR and optical remote sensing images with cascaded-residual adversarial networks[J]. Science China Information Sciences, 2021, 64(2): 1-15.
- [18] Zhao B, Zhang S Z, Xu C X, et al. Deep fake geography? When geospatial data encounter artificial intelligence[J]. Cartography and Geographic Information Science, 2021, 48(4): 338-352.
- [19] Zhang H, Goodfellow I J, Metaxas D N, et al. Self-attention generative adversarial networks[C]// International Conference on Machine Learning, November 19-22, 2019, Long Beach, California, USA. Cambridge: PMLR, 2019: 7354-7363.
- [20] Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis [EB/OL]. (2019-02-25) [2021-08-04]. <https://arxiv.org/abs/1809.11096>.
- [21] Anokhin I, Demochkin K, Khakhulin T, et al. Image generators with conditionally-independent pixel synthesis [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 19-25, 2021, Nashville, Tennessee, USA. New York: IEEE Press, 2021: 14278-14287.
- [22] Xu W J, Long C J, Wang R S, et al. DRB-GAN: a dynamic ResBlock generative adversarial network for artistic style transfer[EB/OL]. (2021-08-17) [2021-08-28]. <https://arxiv.org/abs/2108.07379>.
- [23] Wang Y, Wang D X, Yuan H C, et al. Underwater image enhancement based on pyramidal attention mechanism and GAN[J/OL]. Laser & Optoelectronics Progress: 1-16(2021-2-22) [2021-08-28]. <http://kns.cnki.net/kcms/detail/31.1690.TN.20210220.1710.006.html>.  
王越, 王德兴, 袁红春, 等. 基于金字塔注意力机制和GAN的水下图像增强[J/OL]. 激光与光电子学进展: 1-16(2021-2-22) [2021-08-28]. <http://kns.cnki.net/kcms/detail/31.1690.TN.20210220.1710.006.html>.
- [24] Ouyang J, Shi Q W, Wang X X, et al. Pedestrian trajectory prediction based on GAN and attention mechanism[J]. Laser & Optoelectronics Progress, 2020, 57(14): 141016.  
欧阳俊, 史庆伟, 王馨心, 等. 基于GAN和注意力机制的行人轨迹预测[J]. 激光与光电子学进展, 2020, 57(14): 141016.
- [25] Cui J H, Zhang Y Z, Wang Z, et al. Light-weight object detection networks for embedded platform[J]. Acta Optica Sinica, 2019, 39(4): 0415006.  
崔家华, 张云洲, 王争, 等. 面向嵌入式平台的轻量级目标检测网络[J]. 光学学报, 2019, 39(4): 0415006.