基于 GCI-CycleGAN 风格迁移的跨模态地理定位方法

李清格，杨小冈*，卢瑞涛，王思宇，范继伟，夏　海

(火箭军工程大学 导弹工程学院，陕西 西安 710025)

摘　要：近年来基于视觉的飞行器自主视觉定位技术发展迅速，是飞行器导航制导、态势感知和自主决策的关键技术之一。针对现有跨模态地理定位任务中存在模态差异大、匹配难度大、定位鲁棒性差等问题，提出了一种基于 GCI-CycleGAN 风格迁移的跨模态地理定位方法，通过将风格迁移算法、特征匹配算法和地理定位方法相结合，实现了飞行器跨模态地理定位。首先，获取无人机航拍的正下视实时红外图像和地理位置信息已知的可见光图像；其次，基于生成对抗网络图像风格转换的思想，设计新的生成对抗损失函数，构建并训练了 GCI-CycleGAN 模型，将可见光图像转换为红外图像；然后，利用 SIFT、SURF、ORB、LoFTR、DFM 匹配算法对生成的红外图像与实时红外图像进行匹配；最后，通过透视变换获得实时红外图像中心点在生成图像中的位置，再将该定位点映射到相应的可见光图像上，得到最终的地理定位结果。实验表明，GCI-CycleGAN 相比 CycleGAN 网络可以有效提高图像风格迁移质量，与 DFM 智能匹配算法结合的匹配成功率最高可达 99.48%，比原始跨模态匹配结果提高了 4.73%，平均地理定位误差仅为 1.37 pixel，取得了更加精确、鲁棒的地理定位结果。

关键词：地理定位；　风格迁移；　智能匹配；　跨模态图像；　深度学习

中图分类号：TP391　　**文献标志码：**A　　**DOI：**10.3788/IRLA20220875

0　引　言

近年来，基于视觉图像的飞行器自主视觉定位技术发展迅速。由于单一模态的传感器获取信息有限，而红外图像 (Infrared Radiation Image, IRI) 反映了物体的热辐射信息，不易受外部光线影响，可以在夜间或烟雾环境中有效成像[1−2]。因此利用航拍的实时红外图像和已知地理信息的可见光图像 (Visible Image, VI) 进行匹配，可以获取更加丰富的信息，从而实现飞行器夜间地理定位需求，满足飞行器导航系统的全天时工作需要，具有重要而广泛的应用前景[3]。

传统的跨模态图像匹配方法可以分为基于区域的匹配方法和基于特征的匹配方法。基于区域的匹配方法是利用图像灰度信息进行相似性度量，如互信息 (MI)[4]、归一化互相关 (NCC)[5] 和绝对误差和 (SAD)[6] 等。这类方法虽然原理简单，但不具备实时性，且容易陷入局部最优解。基于特征的匹配方法是通过提取跨模态图像之间的点、线、面等局部不变特征从而实现匹配。典型算法有 SIFT[7]、SURF[8] 等。此类方法虽然计算量小，但提取到的特征较浅，人为定义的特征点无法体现语义信息，容易造成误匹配。

近年来，基于深度学习的匹配方法取得了重大进展，SuperPoint[9] 利用全卷积网络自动提取特征点，SuperGlue[10] 在此基础上利用注意力图神经网络对提取的特征点进行匹配，受此启发，D2-Net[11]、COTR[12] 等算法均利用深度神经网络提取图像特征，并取得了很好的结果。此外，LoFTR[13] 算法基于 Transformer 的自注意力和交叉注意力机制，实现了无检测器的局部特征匹配，在低纹理区域也能取得良好的匹配结果。DFM[14] 算法与 LoFTR 算法都采用了由粗到细的匹配策略，虽然 DFM 与 LoFTR 算法相比匹配点少，但获得了更高的匹配精度。然而，对于可见光和红外跨模态图像，由于图像间存在显著的模态差异，因此基于深度学习的智能匹配方法很难直接应用于跨模

收稿日期：2022−12−06；　**修订日期：**2023−03−28

基金项目：国家自然科学基金项目 (62276274)；航空科学基金项目 (201851U8012)

作者简介：李清格，女，博士生，主要从事视觉导航、目标检测、图像处理等方面的研究。

导师(通讯作者)简介：杨小冈，男，教授，博士生导师，博士，主要从事视觉导航、目标检测、图像处理等方面的研究。

态图像匹配。

针对模态相差较大的跨模态图像匹配任务，利用生成对抗网络 (Generative Adversarial Networks, GAN)[15] 对跨模态图像进行风格迁移后再进行匹配是一种简单且有效的方法[16−17]。图像的转换效果显著影响匹配结果，循环一致性生成对抗网络 (Cycle Consistent Generative Adversarial Networks, CycleGAN)[18] 是典型的风格迁移模型，可以取得很好的转换效果。文献 [19−20] 利用 GAN 网络，成功将光学图像转换为 SAR 等其他模态的图像，可见 GAN 网络在跨模态图像匹配中具有优势和研究意义。

基于以上分析，文中提出了一种基于 GCI-CycleGAN 风格迁移的跨模态图像匹配地理定位方法，如图 1 所示。首先，提出了跨模态地理定位的一般性框架，将特征差异较大的跨模态图像转换到同一特征域内，解决了跨模态图像成像差异带来的误匹配问题；其次，构建了 GCI-CycleGAN 网络对可见光图像进行风格迁移，设计了 Sigmoid 和二值交叉熵损失结合的生成对抗损失函数，并加入了本体映射损失，加快了网络的收敛；然后，利用 SOTA 的智能匹配算法对风格迁移的图像和实时红外图像进行匹配，显著提高了匹配精度；最后，利用透视变换得到地理定位结果。实验结果证明所提算法可以显著提高跨模态图像的匹配性能，实现了可靠、有效的地理定位结果，验证了所提方法的有效性和优越性。

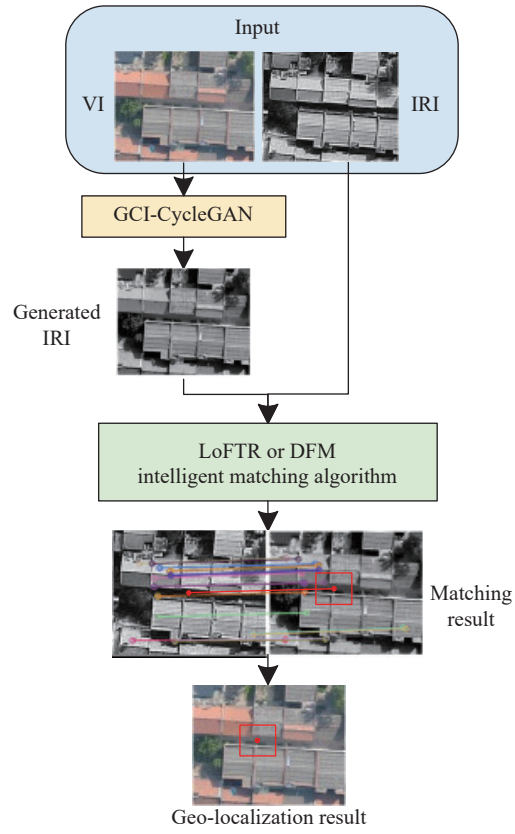

图 1 跨模态地理定位方法框架

Fig.1 Framework of the cross-modal geo-localization method

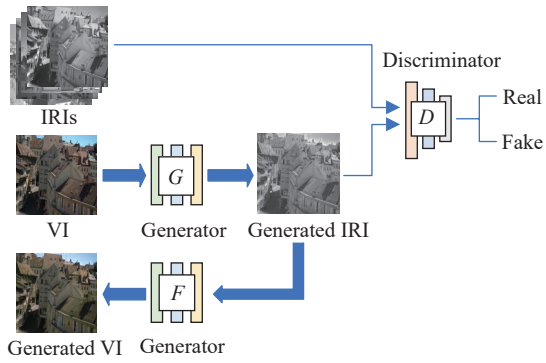

图 2 跨模态图像风格迁移原理

Fig.2 Principle of the cross-modal images style translation

1 GCI-CycleGAN 跨模态迁移模型

1.1 GCI-CycleGAN 结构设计

文中构建 GCI-CycleGAN 模型进行可见光图像到红外图像的风格迁移，通过生成器和判别器之间的对抗训练，学习数据集中红外和可见光图像的像素概率分布来生成图片。此外，通过无限减小生成图像与真实图像之间的差距，实现网络性能优化，转换原理如图 2 所示。

GCI-CycleGAN 的结构与经典 CycleGAN 网络类似，是由两个方向相反的 GAN 网络组成的环形结构，可以实现图像在源域X(可见光图像域) 和目标域Y(红外图像域) 间的相互转换，模型结构如图 3 所示。GCI-CycleGAN 模型主要包括生成器G、F和判别器D_X、D_Y。图中，x、y分别为X域、Y域中的图像。生成器G用于将图像从X域映射到Y域，相反，生成器F用于将Y域图像转换到X域。

以图 3 中的上行分支为例，通过训练两个生成器$G_{X\to Y}$和$F_{Y\to X}$，使得X域图像经过$G:X\to Y$映射后无限逼近Y域中的图像，然后结合判别器D_Y进行对抗性训练，进一步优化图像转换的效果。由于X域和Y域图像不存在一一对应关系，训练过程是随机匹配的，因此 GCI-CycleGAN 中利用循环一致性损失 (cycle-

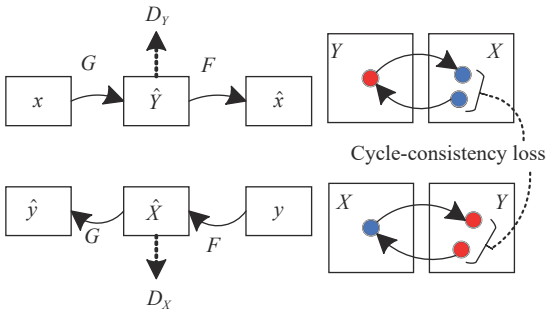

图 3 GCI-CycleGAN 模型结构图

Fig.3 Diagram of the GCI-CycleGAN model structure

consistency loss) 来约束生成器保有源域图像的轮廓信息和内容特征, 防止生成器G、F相互矛盾, 使模型训练更加稳定, 加速了模型的收敛, 增强了多域图像风格转换效果。

1.2 改进损失函数分析

传统 CycleGAN 的损失函数包括两部分: 两个生成对抗损失和循环损失。其中, 循环损失采用$L1$损失函数进行计算, 生成对抗损失使用均方误差 (Mean Square Error, MSE) 进行计算。由于 MSE 的平方运算会放大较大 (> 1) 的误差, 因此离群点会显著影响预测结果, 最终降低模型整体性能。此外, 若初始输出值较大, MSE 损失函数梯度更新幅度较小, 导致收敛时间长, 模型训练不稳定。因此, 文中设计了二值交叉熵 (Binary Cross Entropy, BCE) 损失函数与 Sigmoid 激活函数结合的损失函数来代替 MSE, 以降低 CycleGAN 模型对异常值的灵敏度, 加速网络的收敛, 提高网络的稳定性。此外, 通过设置交叉熵损失的类别权重, 可以缓解样本的不均衡问题。二值交叉熵损失函数的计算公式如下:

$$L_{BCE} = -[x\log\widehat{y} + (1-x)\log(1-\widehat{y})] \tag{1}$$

式中: x为真值; $\widehat{y}$为估计值。此时损失函数的梯度更新与估计值和真值之差成正相关, 模型收敛更快, 提升了数值计算的稳定性。

文中设计的 GCI-CycleGAN 的损失函数包括三部分, 除了原始 CycleGAN 中的生成对抗损失和循环一致性损失外, 还构建了本体一致性损失, 用于维持图像色调, 防止图像整体颜色发生变化。各部分损失函数介绍如下。

1.2.1 生成对抗损失

生成对抗损失是生成器和判别器之间的博弈。GCI-CycleGAN 中将原始的 MSE 生成对抗损失函数替换为 BCE 与 Sigmoid 函数结合的损失函数计算方式, 形成新的生成对抗损失, 包括两部分, 一个是生成器G和判别器D_Y间的生成损失, 如公式 (2) 所示, 另一个是生成器F和判别器D_X间的对抗损失, 如公式 (3) 所示。

$$\begin{aligned} L_{GAN}(G,D_Y,X,Y) =& \mathbb{E}_{y\sim p_Y(y)}[\log D_Y(y)] + \\ & \mathbb{E}_{x\sim p_X(x)}[\log(1-D_Y(G(x)))] \end{aligned} \tag{2}$$

$$\begin{aligned} L_{GAN}(F,D_X,Y,X) =& \mathbb{E}_{x\sim p_{data}(x)}[\log D_X(x)] + \\ & \mathbb{E}_{y\sim p_{data}(y)}[\log(1-D_X(F(y)))] \end{aligned} \tag{3}$$

式中: $\mathbb{E}$为期望值; $p_X(x)$、$p_Y(y)$为实际数据分布。对于公式 (2), $G(x)$为生成器G生成的Y域图像; D_Y用于区分Y域的真实图像y和生成图像$G(x)$。相似的, $F(y)$为生成器F生成的X域图像, D_X用于区分X域的真实图像x和生成图像$F(y)$。

1.2.2 循环一致性损失

循环一致性损失是为了保证输入图像经过图像转换周期后, 能够无限接近原始图像, 即正向循环回路$x \to G(x) \to F(G(x)) \approx x$和反向循环回路$y \to F(y) \to G(F(y)) \approx y$。总的循环一致性损失利用重构图像与真实图像间的$L1$距离进行计算, 描述如下:

$$\begin{aligned} L_{cyc}(G,F) =& \mathbb{E}_{x\sim p_X(x)}[\| F(G(x)) - x\|_1] + \\ & \mathbb{E}_{y\sim p_Y(y)}[\| G(F(y)) - y\|_1] \end{aligned} \tag{4}$$

式中: $F(G(x))$和$G(F(y))$分别表示正向循环回路和反向循环回路的重构图像; $\|*\|_1$表示$L1$范数。

1.2.3 本体一致性损失

GCI-CycleGAN 构建了本体一致性损失用于约束生成器对图像颜色的保持, 防止生成图像色调改变, 确保生成图像保留原始图像的颜色配置。本体一致性损失可以保证图像y送入生成器G, 或图像x送入生成器F后, 输出仍为其本身。本体一致性损失使用$L1$距离计算时, 可描述为:

$$\begin{aligned} L_{identity}(G,F) =& \mathbb{E}_{y\sim p_Y(y)}[\|G(y) - y\|_1] + \\ & \mathbb{E}_{x\sim p_X(x)}[\|F(x) - x\|_1] \end{aligned} \tag{5}$$

1.2.4 目标函数

GCI-CycleGAN 的总损失由上述三类损失构成, 总损失函数表达式为:

$$\begin{aligned} L(G,F,D_X,D_Y) =& L_{GAN}(G,D_Y,X,Y) + L_{GAN}(F,D_X,Y,X) + \\ & \alpha L_{cyc}(G,F) + \beta L_{identity}(G,F) \end{aligned} \tag{6}$$

式中：α、β分别为循环一致性损失和本体一致性损失的权重系数，通过调节α和β可以获取较优的结果。

GCI-CycleGAN 优化的目标函数表达式为：

$$G^*,F^*=\arg\min_{G,F}\max_{D_X,D_Y}L(G,F,D_X,D_Y) \tag{7}$$

1.3 模型训练

为了获得最佳图像转换效果，首先需要对 GCI-CycleGAN 模型进行训练。训练过程采用开源的 RGB-NIR 场景数据集，包含 9 类可见光和近红外图像对，共计 954 张图像。图像输入网络之前，首先归一化为 256×256 大小。归一化后的训练样本示例图像如图 4 所示。

GCI-CycleGAN 模型是在配备有 2 颗 Intel Xeon Gold 6230(2.1 GHz/20 C/27.5 ML3)处理器、2 块NVIDIA RTX 8000 GPU 的浪潮机架式服务器上进行训练。模型训练的损失函数曲线如图 5 所示。

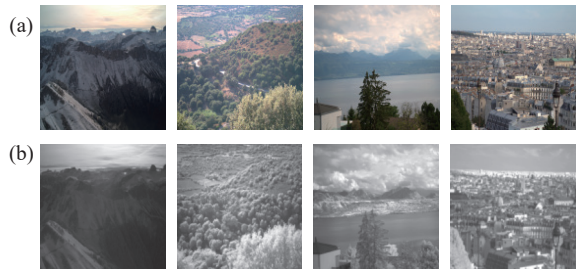

图 4 训练样本示例。(a) 可见光图像；(b) 红外图像

Fig.4 Example of training samples. (a) VIs; (b) IRIs

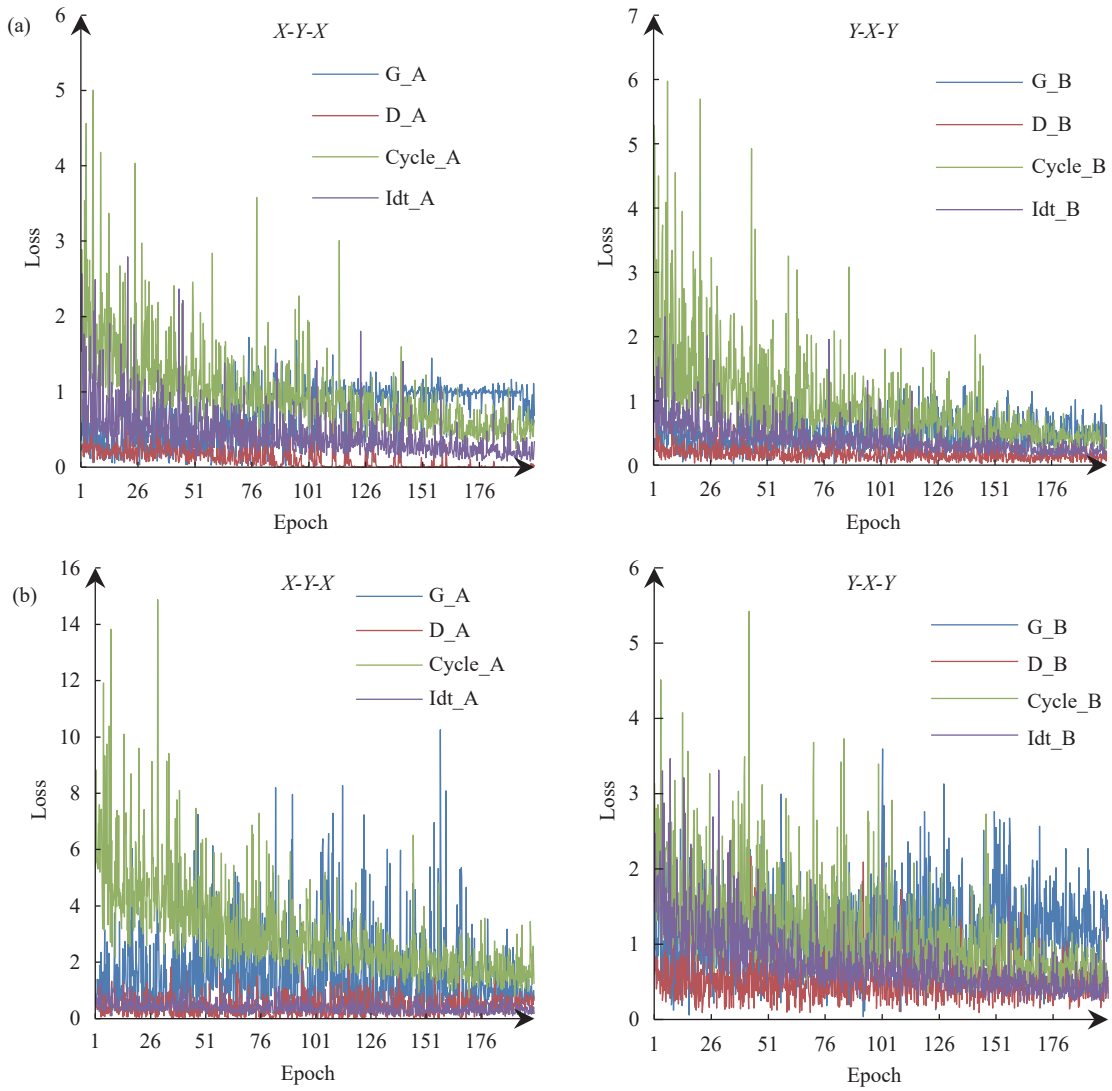

图 5 损失函数曲线。(a) CycleGAN；(b) GCI-CycleGAN

Fig.5 Loss function curve. (a) CycleGAN; (b) GCI-CycleGAN

在 Ubuntu 18.04 操作系统下，模型运行环境基于 CUDA 10.1 和深度学习框架 PyTorch 1.6。模型训练过程中，迭代次数 (epoch) 设置为 200，批次设置为 1。对于前 100 个 epoch，模型学习率设为 0.000 2，后 100 个 epoch 利用 Adam 优化器自适应地调整学习率。为了增强可见光图像到红外图像的转换效果，GCI-CycleGAN 模型中将$X \to Y \to X$域的正向重建损失权重设置为 30，$Y \to X \to Y$域的反向重建损失权重设置为 10，强化了生成器$G_{X \to Y}$的重要度。此外，循环一致性损失L_{cyc}与本体一致性损失$L_{identity}$的权重系数之比$\alpha:\beta$设置为3∶1。

由图 5 可知，随着模型训练的进行，损失函数总体趋势在不断减小，但是存在上下波动的情况，证明了生成器和判别器在不断地进行博弈，特别是 GCI-CycleGAN 的损失函数曲线波动更为剧烈，说明 GCI-CycleGAN 中生成器和判别器的博弈过程比原始 CycleGAN 更加强烈。此外，由于提高了正向重建损失权重，可以看到 GCI-CycleGAN 中G_A损失的最终训练结果显著优于 GCI-CycleGAN 中的G_B损失和 CycleGAN 的G_A损失。

2 智能匹配定位算法

2.1 LoFTR

2.1.1 模型架构

LoFTR 是一种端到端的智能匹配方法，与传统先检测再匹配的方法相比，LoFTR 无需先检测角点等纹理清晰的点，利用 Transformer 的全局感受野特性，可以在低纹理区域实现密集匹配，其模型架构如图 6 所示。

其中，图像 A 为实时红外图像，图像 B 为生成的红外图像。LoFTR 采用“先粗后细”的匹配方式，首先利用 ResNet-18 和 FPN 网络提取1/8和1/2尺度的特征图，分别用于粗匹配和精匹配。1/8尺度的特征图经过展平操作和位置编码后，利用 Transformer 中的自注意力和交叉注意力机制提取特征图中的上下文信息，得到较易匹配的特征图，然后依据匹配策略在低分辨率特征图上进行粗略级像素密集匹配。粗匹配结果映射到1/2尺度的特征图上，然后在高分辨率特征图上进行亚像素级精细匹配，优化匹配结果。

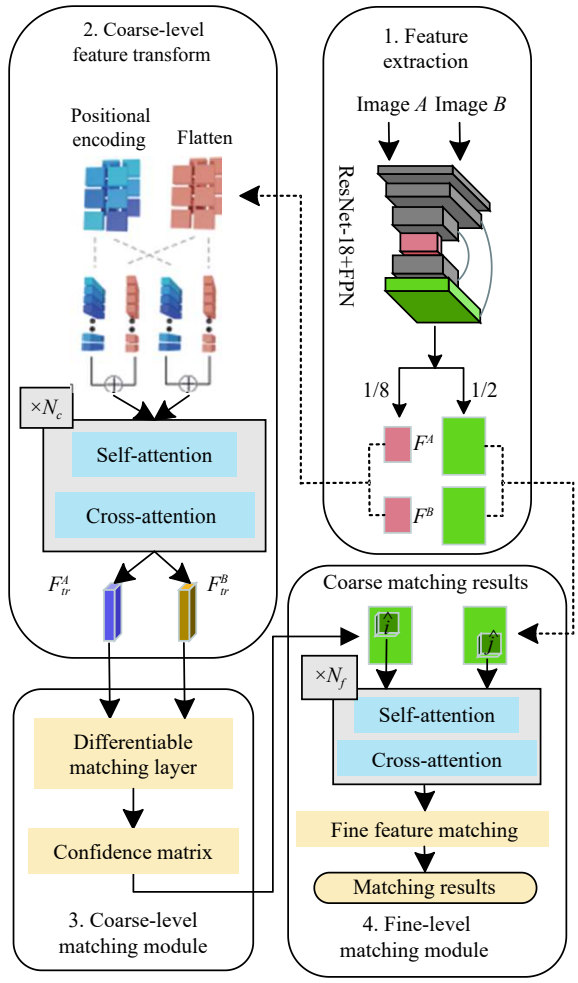

图 6 LoFTR 模型结构图

Fig.6 Diagram of the LoFTR model structure

2.1.2 匹配策略

(1) 粗粒度匹配

LoFTR 的粗匹配过程首先是计算特征图F^A_{tr}和F^B_{tr}之间的匹配得分矩阵S，过程如下：

$$S(i,j)=\frac{1}{\tau}\cdot\left\langle F^A_{tr}(i),F^B_{tr}(j),\right\rangle \tag{8}$$

其次利用最优传输算法或 dual-softmax 方法计算最优匹配，利用 dual-softmax 方法可以得到置信度矩阵 (confidence matrix) P_c：

$$P_c(i,j)=\mathrm{softmax}(S(i,\cdot))_j\cdot\mathrm{softmax}(S(\cdot,j))_i \tag{9}$$

然后根据置信阈值θ_c得到匹配结果。最后利用最近邻 (Mutual Nearest Neighbor, MNN) 算法滤除掉异常匹配，得到粗匹配预测M_c：

$$M_c = \left\{(\widetilde{i},\widetilde{j}) \mid \forall(\widetilde{i},\widetilde{j}) \in \mathrm{MNN}(P_c), P_c(\widetilde{i},\widetilde{j}) \geqslant \theta_c\right\} \tag{10}$$

(b) 细粒度匹配

精匹配过程首先是将粗匹配得到的每对点对$(\widetilde{i},\widetilde{j})$映射到对应的细粒度特征图 (1/2尺度) 上，然后利用$w\times w$的窗口对特征图进行裁剪，然后将局部窗口输入至自注意力和交叉注意力机制进行特征提取，得到以$\widetilde{i}$、$\widetilde{j}$为中心的特征图$\widehat{F}^A_{tr}(\widetilde{i})$和$\widehat{F}^B_{tr}(\widetilde{j})$。然后计算$\widehat{F}^A_{tr}(\widetilde{i})$中心特征与$\widehat{F}^B_{tr}(\widetilde{j})$中所有特征的匹配概率，通过计算匹配概率分布，即可得到$\widehat{F}^B_{tr}(\widetilde{j})$中的亚像素级别的匹配点$\hat{j}'$。计算所有的匹配点对$\left\{(\hat{i},\hat{j}')\right\}$，从而得到最终的精匹配预测$M_f$。

2.1.3 损失函数

LoFTR 的总损失函数包括粗粒度损失和细粒度损失这两个部分。粗粒度损失函数采用置信矩阵P_c的负对数似然损失，计算如下：

$$L_c = -\frac{1}{\left|M_c^{gt}\right|}\sum_{(\hat{i},\hat{j})\in M_c^{gt}} \log P_c(\hat{i},\hat{j}) \tag{11}$$

式中：M_c^{gt}为粗匹配的真值，可通过数据集中的相机位姿和深度信息计算得到。

细粒度损失采用$L2$损失进行计算，通过计算每个特征点$\widetilde{i}$与匹配点$\hat{j}'$的距离得到的位置误差：

$$L_f = \frac{1}{\left|M_f\right|}\sum_{(\hat{i},\hat{j}')\in M_f}\frac{1}{\sigma^2(\hat{i})}\left\|\hat{j}' - \hat{j}'_{gt}\right\|_2 \tag{12}$$

式中：$\sigma^2(\hat{i})$为每个特征点$\widetilde{i}$生成热力图的方差；$\hat{j}'_{gt}$是通过计算相机位姿和深度信息从$\widetilde{i}$空间变换到$\widehat{F}^B_{tr}(\widetilde{j})$得到的，若变换后超出$w\times w$窗口的范围，则舍弃。

综上所述，总损失函数为：

$$L_{LoFTR} = L_c + L_f \tag{13}$$

2.2 DFM 算法架构及原理

DFM 是一种两阶段智能匹配算法，利用深度神经网络提取图像特征后，通过粗略的几何变换估计实现图像配准，然后采用从粗到细的策略优化特征点的匹配结果，模型结构见图 7。

如图所示，DFM 算法首先利用 VGG-19 深度神经网络对输入的图像进行特征提取，充分获取图像的深度特征信息。在第一阶段，采用密集最近邻搜索策略 (Dense Nearest Neighbor Search, DNNS) 在低空间分辨率下进行计算，得到待匹配图像之间几何变换的粗略估计。在第二阶段，首先根据第一阶段的粗略估计值对生成的红外图像进行几何变换，然后使用实时红外图像特征图和几何变换后的生成红外图像最后几层特征图利用 DNNS 进行计算，实现更高分辨率上的层次化匹配优化。

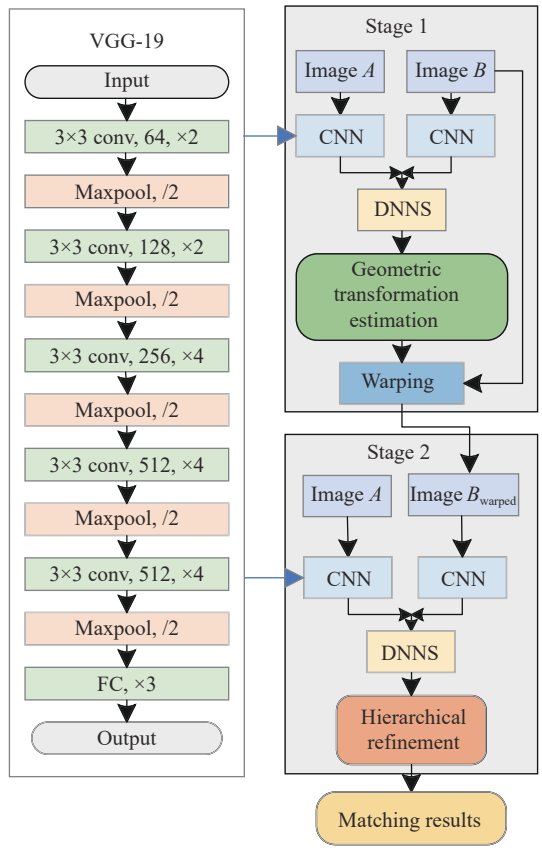

图 7 DFM 模型结构图

Fig.7 Diagram of the DFM model structure

3 实验结果与分析

3.1 跨模态图像转换实验

跨模态图像转换实验利用已成功训练的 CycleGAN 模型和 GCI-CycleGAN 模型将可见光图像转换为红外图像，测试数据采用 42 对可见光和红外无人机航拍图像对，大小为 256×256。部分实验结果如图 8 所示。

从实验结果可以看出，生成的红外图像与真实红外图像相比，具有模态一致性，且 CycleGAN 和 GCI-CycleGAN 在改变可见光图像模态的同时，保证了生成的红外图像的尺寸、结构和视场不会发生变化，仅在模态上无限接近真实的红外图像。此外，与 CycleGAN

图 8 (a) 待转换的可见光图像; (b) CycleGAN 转换的红外图像; (c) GCI-CycleGAN 转换的红外图像; (d) 真实的红外图像

Fig.8 (a) VIs to be converted; (b) IRIs converted by CycleGAN; (c) IRIs converted by GCI-CycleGAN; (d) Real IRIs

相比, GCI-CycleGAN 生成的红外图像在亮度、对比度上更接近目标红外图像, 且 GCI-CycleGAN 更注重细节纹理特征的表达, 无失真和畸变现象, 具有更好的风格迁移效果。

此外, 为了评估不同风格迁移算法的性能, 文中选用峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR)、平均哈希 (average Hash, aHash)、感知哈希 (perceptual Hash, pHash) 和学习感知图像块相似度 (Learning Perceptual Image Patch Similarity, LPIPS)[21] 算法计算图像的相似度, 从而对可见光图像到红外图像的转换结果进行评估。其中, PSNR 越大, aHash、pHash 和 LPIPS 越小, 代表图像相似度越高。

通过计算测试集中的实时红外图像与可见光图像和风格迁移后的红外图像在上述 4 种评估算法上的平均值, 对 GCI-CycleGAN 和 CycleGAN 模型的图像转换效果进行定量分析, 计算结果如表 1 所示。

表中数据分别为真实红外图像与可见光图像、CycleGAN 和 GCI-CycleGAN 模型转换后的红外图像在 PSNR、aHash、pHash 和 LPIPS 指标上的计算结果。可以看到, 与可见光图像相比, 转换后的红外图像与实时红外图像的相似程度更高, 特别是 GCI-CycleGAN 生成的红外图像与实时红外图像相似度最高, 可以证明 GCI-CycleGAN 算法的有效性。

表 1 不同模型性能对比

Tab.1 Performance comparison of different models

	PSNR/dB	aHash	pHash	LPIPS
Original images	14.771	12.400	10.200	0.342
CycleGAN	19.176	11.200	10.000	0.208
GCI-CycleGAN	**19.422**	**10.400**	**9.800**	**0.191**

3.2 智能匹配定位实验

为了验证 GCI-CycleGAN 图像转换算法作为跨模态图像匹配的预处理方法的可靠性和有效性, 采用不同的匹配算法来评估图像匹配效果。图像匹配实验的测试集一部分与 3.1 节图像转换实验中的测试集相同, 大小为 256×256 的可见光和近红外航拍图像对, 记为 TS1。另一部分是 CycleGAN 和 GCI-CycleGAN 对 TS1 中的可见光进行图像风格迁移, 得到的生成红外图像与实时红外图像组成的测试数据集, 分别记为 TS2 和 TS3。智能匹配算法 (LoFTR、DFM) 和传统匹配算法 (SIFT、SURF、ORB) 在 TS1、TS2 和 TS3 上的匹配结果如图 9 所示。其中, 传统匹配算法在经过特征点提取、特征描述和特征点匹配后, 采用 GMS 算法对匹配结果进行优化, 以筛除误匹配点。

从图 9 可以看出, 对于同一种匹配算法来说, 可见光和实时红外图像 (TS1) 直接匹配的特征点数量较少, 且存在一定的误匹配, 匹配效果较差。与之相比, GCI-CycleGAN 和 CycleGAN 转换后的红外图像与实时红外图像匹配的特征点明显增多, 且误匹配点较少。特别是 GCI-CycleGAN 转换后的红外图像与实时红外图像 (TS3) 成功匹配的特征点最多。

对比不同匹配算法在同一测试集上的匹配结果可以看出, 传统匹配方法 SIFT、SURF 和 ORB 成功匹配到的特征点数量远远小于 LoFTR 和 DFM 智能算法的匹配结果。此外, GCI-CycleGAN 算法和传统匹配算法结合的匹配结果虽然有所提升, 但提升效果并不显著, 其匹配成功的特征点数量与直接采用智能匹配算法的结果相当。原因在于 GCI-CycleGAN 对图像进行风格迁移后, 虽然减小了跨模态图像间的模态差异, 但转换后的红外图像与真实红外图像之间仍为异源图像, 且图像在分辨率和细节纹理方

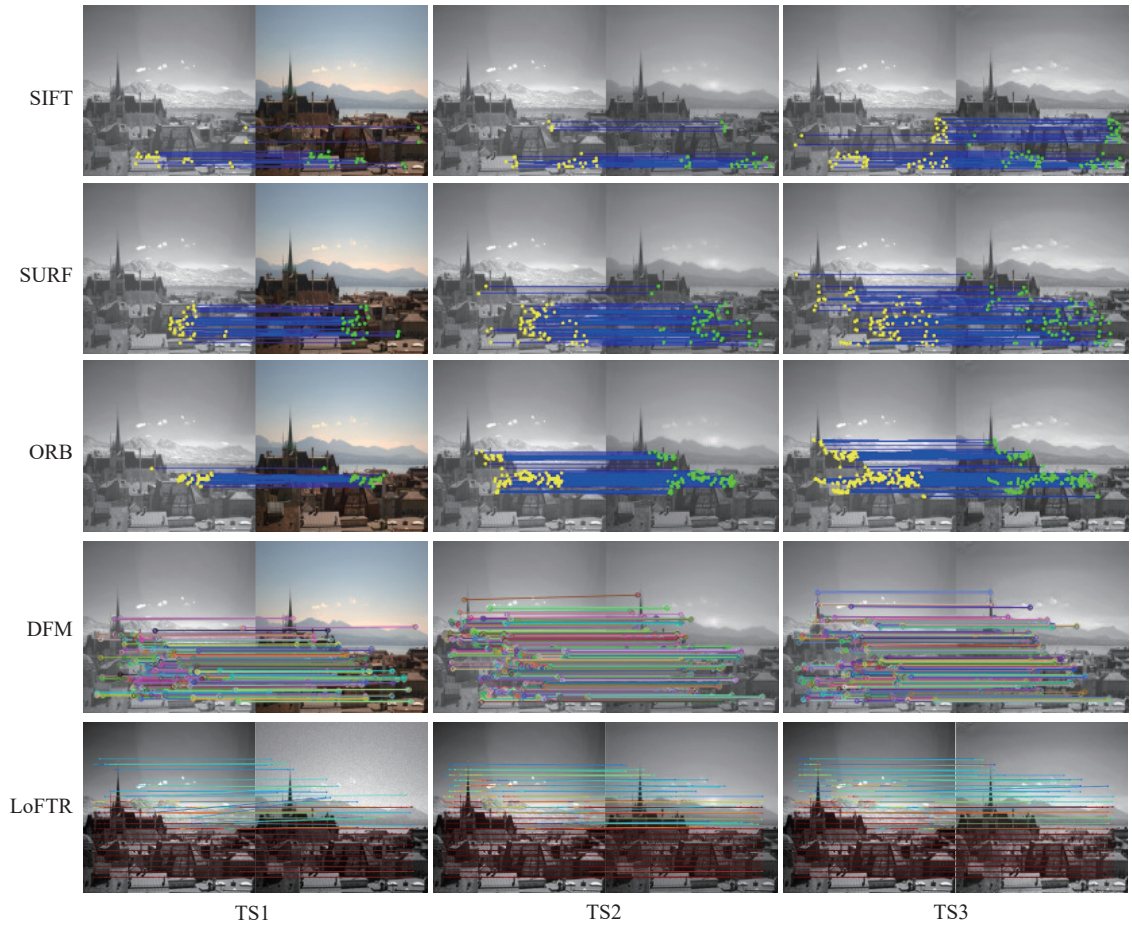

图 9 不同匹配方法结果对比

Fig.9 Comparison of the matching methods results

面仍存在一定的差异。而基于深度学习的智能匹配算法匹配到的特征点显著增加，匹配结果也更加密集，特别是对于特征点较少的稀疏纹理区域也可以实现匹配。

为了定量分析匹配实验结果，采用正确匹配点 (Correct Matching Points, CMP) 和匹配成功率 (Match Success Rate, MSR) 两种评价指标对匹配算法进行评估。正确匹配点是满足公式 (14) 的所有匹配点对的数量，反映了不同匹配方法对于各类输入图像的匹配能力。

$$\mathrm{CMP}=\sqrt{(x_i-x_i')^2+(y_i-y_i')^2} \tag{14}$$

匹配成功率是指正确匹配上的总特征点占总特征点数的比例，计算如下：

$$\mathrm{MSR}=\frac{m2}{p_A+p_B} \tag{15}$$

式中：m为正确匹配点对数；p_A为实时红外图像中的特征点数量；p_B为待匹配的可见光或转换后的红外图像中提取到的特征点数量。正确匹配率越高，则代表算法的正确匹配上的特征点越多，匹配性能则越好。不同的匹配算法在 TS1、TS2 和 TS3 上的实验结果如表 2 所示。

从实验结果可以看出，对于同一种匹配算法，经 GCI-CycleGAN 转换过的红外图像提取到的特征点数量更多，正确匹配点的数量和匹配成功率显著增加，证明 GCI-CycleGAN 风格迁移后再进行匹配的可靠性和有效性。对于 SIFT、SURF 和 ORB 传统匹配算法来说，由于 SIFT 匹配算法提取的特征点数和匹配成功的特征点数较少，因此实时性更高。与传统匹配算法对比，LoFTR 和 DFM 智能算法的实时性略低，但正确匹配点的数量更加显著，匹配成功率有大幅提升，体现了智能匹配算法的优越性和必要性。

表 2 匹配方法性能对比

Tab.2 Performance comparison of matching methods

Methods	Dataset	p_A	p_B	CMP	MSR	FPS/frame·s^{-1}
SIFT+GMS	TS1	353	336	24	6.97%	25
	TS2	352	181	29	10.88%	2.33
	TS3	353	359	63	17.70%	2.38
SURF+GMS	TS1	350	350	38	10.86%	3.03
	TS2	350	333	62	18.16%	1.39
	TS3	350	357	88	24.89%	1.63
ORB+GMS	TS1	466	453	58	12.62%	3.23
	TS2	466	453	159	34.60%	1.45
	TS3	466	455	257	55.81%	1.43
LoFTR	TS1	385	385	301	78.18%	2.33
	TS2	437	437	411	94.05%	1.23
	TS3	442	442	432	97.74%	1.25
DFM	TS1	305	305	289	94.75%	1.52
	TS2	434	434	428	98.62%	0.95
	TS3	**581**	**581**	**578**	**99.48%**	0.96

3.3 跨模态地理定位实验

实验利用道通 EVO Ⅱ 无人机在指定区域按规划路径进行巡航，飞行高度为 350 m，拍摄得到实时红外图像和已知地理位置信息的可见光图像，如图 10 所示。实验图像中包含房屋、道路、植物等，视角为正下视。图 10 (a) 为地理位置信息已知的可见光图像，图 10 (b) 为无人机拍摄的实时红外图像，图 10 (c) 为经过 GCI-CycleGAN 模型风格迁移后的红外图像。

图 10 地理定位数据集图像示例。(a) 可见光图像；(b) 实时红外图像；(c) 生成红外图像

Fig.10 Example images of the geo-location dataset. (a) Visible images; (b) Real-time infrared images; (c) Generated infrared images

文中实验首先将地理信息已知的可见光图像通过 GCI-CycleGAN 模型进行风格迁移，使其从可见光图像域转换至红外图像域，同时保持细节纹理特征不变。然后将风格迁移后的红外图像与实时航拍红外图像进行智能匹配，根据特征点匹配关系计算单应性变换矩阵，再通过透视变换求出实时红外图像中心点 (飞行器当前位置) 在生成红外图像中对应的像素坐标。最后将风格迁移图像中求得的定位点映射到可见光图像上。由于可见光图像中各像素点对应的地理位置信息已知，因此可以得到最终的无人机地理定位结果。

图11 对比了基于SIFT、SURF、ORB、DFM 和LoFTR 五种匹配方法的地理定位效果，可以看出，匹配算法的性能显著影响最终的地理定位结果。不同方法的定位点坐标和平均误差如表 3 所示，可知智能匹配算法的定位误差远小于传统算法，特别是 DFM 算法的平均定位误差仅为 1.37 pixel。

综上所述，将无人机实际飞行轨迹与匹配定位结果作对比，如图 12 所示。从实验结果可以直观看出，文中提出的基于 GCI-CycleGAN 风格迁移的跨模态地理定位方法可以取得良好的定位效果，定位结果与真实无人机飞行轨迹基本一致，定位点位置偏差较小，验证了文中所提地理定位方法具有有效性和可靠性，可以用于飞行器视觉导航任务。

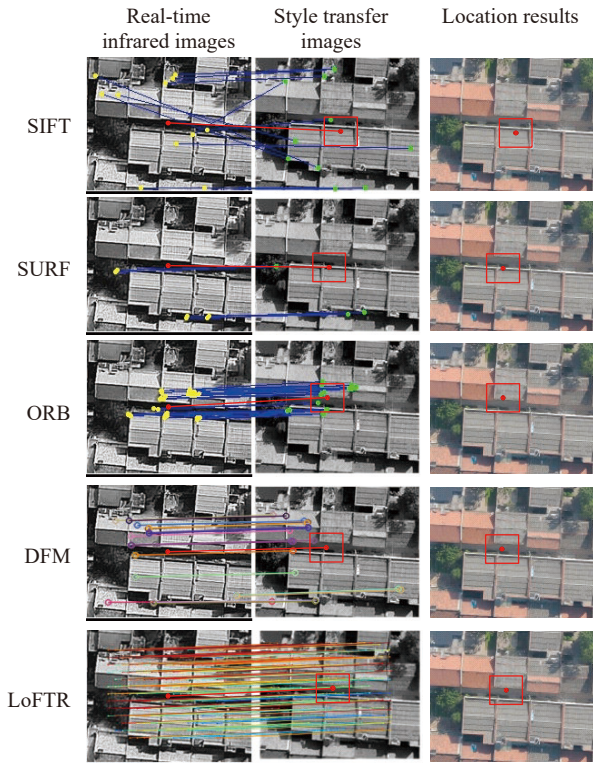

图 11 匹配和地理定位结果

Fig.11 Diagram of the matching and geo-location results

表 3 地理定位性能

Tab.3 The performance of geo-location

	Group	True location	SIFT	SURF	ORB	LoFTR	DFM
Positioning coordinates	1	(52, 43)	(43, 42)	(40, 39)	(50, 34)	(51, 42)	(52, 44)
	2	(46, 36)	(54, 45)	(47, 42)	(45, 33)	(48, 36)	(45, 37)
	3	(54, 38)	(51, 37)	(48, 32)	(48, 35)	(52, 39)	(54, 40)
	4	(48, 40)	(50, 43)	(48, 45)	(51, 48)	(48, 38)	(49, 40)
	5	(48, 34)	(42, 35)	(40, 38)	(42, 37)	(46, 35)	(48, 35)
	6	(46, 38)	(50, 40)	(47, 43)	(52, 41)	(46, 39)	(47, 39)
Average errors/pixel		-	6.52	8.30	6.84	1.81	**1.37**

图 12 实际飞行轨迹与定位结果对比

Fig.12 Comparison between actual flight trajectory and location results

4 结 论

文中通过对飞行器航拍红外图像与可见光图像的风格迁移，研究了跨模态图像匹配的地理定位问题。提出了一种基于 GCI-CycleGAN 的跨模态图像智能匹配方法，将生成对抗网络与匹配算法结合，来解决基于可见光和红外航拍图像匹配的地理定位问题。首先通过设计新的损失函数构建了 GCI-CycleGAN 模型对可见光图像进行风格迁移，然后利用 LoFTR 和 DFM 智能匹配算法实现生成图像与实时红外图像的有效匹配，最后将匹配关系映射到原始跨模态图像对上，得到最终的地理定位结果。实验结果表明，文中方法有效实现了图像的跨模态转换，显著提高了匹配算法的成功匹配率，证明了该地理定位方法的价值和意义。在未来，如何在嵌入式边缘计算设备中部署文中所提算法，同时平衡成本、功耗和算力，使算法满足有效性和实时性，这在当前实际工程应用中是一个具有挑战性的难题。

参考文献：

[1] Lu R, Shen T, Yang X, et al. Infrared small moving target detection algorithm based on incremental inertial navigation information in high dynamic air to ground background (Invited) [J]. *Infrared and Laser Engineering*, 2022, 51(4): 20220191. (in Chinese)

[2] Gao F, Yang X, Lu R, et al. Anchor-free lightweight infrared object detection method (Invited) [J]. *Infrared and Laser Engineering*, 2022, 51(4): 20220193. (in Chinese)

[3] Yang X, Cheng S, Xi J. Aircraft Heterogeneous Scene Matching

Guidance Technology[M]. Beijing: Science Press, 2016. (in Chinese)

[4] Suri S, Reinartz P. Mutual-information-based registration of terraSAR-X and ikonos imagery in urban areas [J]. ***IEEE Transactions on Geoscience and Remote Sensing***, 2010, 48(2): 939-949.

[5] Fang Y, Hu J, Du C, et al. SAR-optical image matching by integrating Siamese U-Net with FFT correlation [J]. ***IEEE Geoscience and Remote Sensing Letters***, 2021, 19: 1-5.

[6] Kwon O. Similarity measures for object matching in computer vision[D]. England: University of Bolton, 2016.

[7] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. ***International Journal of Computer Vision***, 2004, 60(2): 91-110.

[8] Bay H, Tuytelaars T, Gool L V. SURF: Speeded up robust features[C]//European Conference on Computer Vision, 2006: 407-417.

[9] Detone D, Malisiewicz T, Rabinovich A. Superpoint: Self-supervised interest point detection and description[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018: 224-236.

[10] Sarlin P-E, Detone D, Malisiewicz T, et al. Superglue: Learning feature matching with graph neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 4938-4947.

[11] Dusmanu M, Rocco I, Pajdla T, et al. D2-net: A trainable cnn for joint description and detection of local features[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 8092-8101.

[12] Jiang W, Trulls E, Hosang J, et al. Cotr: Correspondence transformer for matching across images[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 6207-6217.

[13] Sun J, Shen Z, Wang Y, et al. LoFTR: Detector-free local feature matching with transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8922-8931.

[14] Efe U, Ince K G, Alatan A. Dfm: A performance baseline for deep feature matching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 4284-4293.

[15] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems, 2014: 139-144.

[16] Chen T, Guo J, Han X, et al. Visible and infrared image matching method based on generative adversarial model [J]. ***Journal of Zhejiang University (Engineering Science)***, 2022, 56(1): 63-64. (in Chinese)

[17] Zhao J, Yang D, Li Y, et al. Intelligent matching method for heterogeneous remote sensing images based on style transfer [J]. ***IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing***, 2022, 15: 6723-6731.

[18] Zhu J, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 2223-2232.

[19] Li X, Du Z, Huang Y, et al. A deep translation (GAN) based change detection network for optical and SAR remote sensing images [J]. ***ISPRS Journal of Photogrammetry and Remote Sensing***, 2021, 179: 14-34.

[20] Song J, Li J, Chen H, et al. MapGen-GAN: A fast translator for remote sensing image to map via unsupervised adversarial learning [J]. ***IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing***, 2021, 14: 2341-2357.

[21] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 586-595.

Cross-modal geo-localization method based on GCI-CycleGAN style translation

Li Qingge，Yang Xiaogang*，Lu Ruitao，Wang Siyu，Fan Jiwei，Xia Hai

(Missile Engineering Institute, PLA Rocket Force University of Engineering, Xi'an 710025, China)

Abstract:

Objective The purpose of this research is to propose a cross-modal image geo-localization method based on GCI-CycleGAN style translation for vision-based autonomous visual geo-localization technology in aircraft. The technology is essential for navigation, guidance, situational awareness, and autonomous decision-making. However, existing cross-modal geo-localization tasks have issues such as significant modal differences, complex matching, and poor localization robustness. Therefore, real-time infrared images and visible images with known geo-location information are acquired with the proposed method, and a GCI-CycleGAN model is trained to convert visible images into infrared images using generative adversarial network image style translation. The generated infrared images are matched with real-time infrared images using various matching algorithms, and the position of the real-time infrared image center point in the generated image is obtained through perspective transformation. The positioning point is then mapped to the corresponding visible image to obtain the final geo-localization results. The research is crucial as it provides a solution to the challenges faced by existing cross-modal geo-localization tasks, improving the quality and robustness of geo-localization outcomes. A higher matching success rate and a more accurate average geo-localization error are achieved with the GCI-CycleGAN and DFM intelligent matching algorithms. The proposed method has significant practical implications for vision-based autonomous visual geo-localization technology in aircraft, which plays a crucial role in navigation and guidance, situational awareness, and autonomous decision-making.

Methods The research describes a proposed method for cross-modal image geo-localization based on GCI-CycleGAN style translation (Fig.1). First, the real-time infrared and visible light images of the drone's direct down view aerial photography are obtained (Fig.10). The GCI-CycleGAN model structure (Fig.3) and the generated confrontation loss function were designed and trained on the RGB-NIR scene dataset (Fig.5). The trained GCI-CycleGAN model is utilized to perform style transfer on visible light images, resulting in more realistic pseudo infrared images (Fig.8). Using various matching algorithms, including SIFT, SURF, ORB, LoFTR (Fig.6), and DFM (Fig.7), the generated pseudo infrared image is matched with the real-time infrared image to obtain the feature point matching relationship (Fig.9). The homography transformation matrix is determined based on the matching relationship of feature points. Based on the homography transformation matrix, perspective transformation is performed on the center point of the real-time infrared image to determine the pixel points corresponding to the center point in the pseudo infrared image. Then the pixel points corresponding to the center point in the pseudo infrared image are mapped to the visible light image, and the mapping points in the visible light image are determined (Fig.11). Finally, based on the geographic location information corresponding to the mapping points in the visible light image, the geographic positioning results of the drone are obtained (Fig.12).

Results and Discussions The experiment results demonstrate that compared to CycleGAN, GCI-CycleGAN pays more attention to the expression of detailed texture features, generates infrared images without distortion, and is closer to the target infrared image in brightness and contrast, effectively improving the quality of image style translation (Tab.1). The combination of GCI-CycleGAN and DFM intelligent matching algorithm can achieve a matching success rate of up to 99.48%, 4.73% higher than the original cross-modal matching result, and the average geo-localization error is only 1.37 pixel, achieving more accurate and robust geo-localization outcome.

Conclusions This article studies the geographic positioning problem of cross-modal image matching through style translation between infrared and visible light images captured by aircraft aerial photography. A cross-modal image intelligent matching method based on GCI-CycleGAN is proposed, which combines generative adversarial networks with matching algorithms to solve the geographic positioning problem based on visible light and infrared aerial image matching. First, a new loss function is designed to construct a GCI-CycleGAN model to transfer the style of visible images, and then LoFTR and DFM intelligent matching algorithms are used to achieve effective matching between the generated image and real-time infrared images. Finally, the matching relationship is mapped to the original cross-modal image pair to obtain the final geographical positioning result. The experimental results show that the proposed method effectively achieves cross-modal transformation of images and significantly improves the success rate of matching algorithms, demonstrating the value and significance of this geographic positioning method. In the future, how to deploy the proposed algorithm in embedded edge computing devices and balance cost, power consumption, and computing power to make the algorithm meet the effectiveness and real-time is a challenging problem in current practical engineering applications.

Key words: geo-localization; style translation; intelligent matching; cross-modal images; deep learning

Funding projects: National Natural Science Foundation of China (61806209); Chinese Aeronautical Establishment (201851U8012)