基于两阶段模型的无人机图像厚云区域内容生成

韦 哲[1)] 李从利[1)] 沈延安[2)] 刘永峰[1)] 周浦城[3)]

[1)](陆军炮兵防空兵学院兵器工程系 合肥 230031)

[2)](陆军炮兵防空兵学院无人机应用系 合肥 230031)

[3)](陆军炮兵防空兵学院信息工程系 合肥 230031)

摘 要 无人机飞行时常因云层遮挡而造成拍摄图像下垫面的信息损失,而现有多光谱和多时相的云区内容估计方法主要针对卫星遥感图像,无法直接应用于无人机图像. 如何利用已有信息合理地推断厚云遮挡区域内容,以提高影像的可用性是亟待解决的问题. 针对无人机成像光谱单一、航时短且航迹随机的特点,利用图像修复理论,提出了一种基于深度卷积生成对抗网络(Deep Convolutional Generative Adversarial Networks, DCGAN)的两阶段厚云区域内容生成方法. 采用词袋(Bag of Words, BoW)检索算法搜索一阶段修复图像的相似样本,设计了仿射网络对它们进行注意力对齐,使生成的图像能利用同分布已知图像的信息,解决了无人机图像含有多种分布而难以有效提取特征,以及现有修复方法强依赖于单幅图像的局限. 改进了经典 DCGAN 的结构,综合局部和全局对抗损失,以及感知损失和总变差损失,设计了新的联合损失函数. 在仿真实验的中心 1/4 掩膜上,峰值信噪比(Peak Signal to Noise Ratio, PSNR)、结构相似性(Structural Similarity, SSIM)、平均像素 L_1 损失、自然图像质量评价 NIQE(Natural Image Quality Evaluator)和 BLIINDS(Blind Image Integrity Notator using DCT Statistics)指标比对比方法分别改善 0.3214~3.6793、0.0005~0.0543、0.0171~4.1120、0.0565~4.7440 和 0.8841~4.2586. 在真实含云图像实验中,NIQE 和 BLIINDS 指标比对比方法分别改善 0.1062~1.8992 和 1.0903~5.6495. 实验结果的主观和客观分析表明,相比经典方法,本文方法在语义合理性、信息准确性和视觉自然性上都具有一定的优势,为厚云遮挡下的单光谱图像像素值预测提供了一种较好的解决方案.

关键词 无人机图像;图像生成;两阶段模型;深度学习;深度卷积生成对抗网络

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2021.02233

Thick Cloud Region Content Generation of UAV Image Based on Two-Stage Model

WEI Zhe[1)] LI Cong-Li[1)] SHEN Yan-An[2)] LIU Yong-Feng[1)] ZHOU Pu-Cheng[3)]

[1)](*Department of Arms Engineering, PLA Army Academy of Artillery and Air Defense, Hefei* 230031)

[2)](*Department of UAV application, PLA Army Academy of Artillery and Air Defense, Hefei* 230031)

[3)](*Department of Information Engineering, PLA Army Academy of Artillery and Air Defense, Hefei* 230031)

Abstract Cloud covering often causes the loss of information on the underlying surface of the image during UAV flight. However, existing cloud region estimating methods based on multi-spectral and multi-temporal mainly orient to satellite remote sensing images, and cannot be applied directly to UAV images. How to use the available information to reasonably infer the content covered by the thick cloud, so as to improve the image availability, remains an urgent problem to be solved. With image inpainting theory, which regards the covered regions as the missing or damaged parts of the image and devotes to reconstruct

收稿日期:2019-08-21;在线发布日期:2020-08-26. 本课题得到安徽省自然科学基金“低照度环境下视频监控图像增强方法研究”(1908085MF208)资助.**韦 哲**,硕士,讲师,主要研究领域为深度学习、图像修复及质量评价. E-mail: mrweichengjin@163.com. **李从利**(**通信作者**),硕士,教授,主要研究领域为图像信息智能处理. E-mail:lcliqa@163.com. **沈延安**,博士,教授,主要研究领域为无人机系统工程. **刘永峰**,博士,副教授,主要研究领域为遥感图像处理. **周浦城**,博士,副教授,主要研究领域为图像处理与分析.

their consistency, a two-stage thick cloud region content generating method based on DCGAN is proposed for the characteristics of single spectrum, short flight time and random flight path of UAV imaging. The two-stage model consists of a first stage DCGAN, an image retrieval module, an affine transformation net and a second stage DCGAN from front to back sequentially. The first stage DCGAN takes the masked image in, and generates preliminary completion result. In order to make the most of the homogeneous samples in dataset, an image retrieval module and an affine transformation is added. BoW retrieval algorithm is used to search for top N homogeneous samples of the image completed in the first stage, and the affine transform network is designed to align them with attention mechanisms for the second stage. The second stage DCGAN, which has the same structure as the first, takes the preliminary completion result and the output of the affine transform network, and generates the refined result in the end. The 4 parts constitute a complete forward form. This model makes image generation easier to utilize the information of the known image with identical distribution, and solves the difficulty of feature extraction with multiple distributions in UAV images, and addresses the limitation that existing inpainting methods rely heavily on single image. This paper also improves the structure of classical DCGAN, and designs a new joint loss function, combining local and global adversarial loss with the perceptual loss and the total variation loss, which not only prevent blurry result, but also generate pixels that approximate the true semantic distribution with less noise. In the training phase, image retrieval module is trained firstly to get the whole bags of visual word and clusters. Then affine transform network is trained by affined samples with manual random setting. 2-stage DCGANs are trained end-to-end using Adam optimization alternately with samples generated by the first 2 module. In the testing phase, these modules are cascaded and worked at fixed parameters. Simulation experiments with masks and real cloud-containing image are carried out respectively. On the central 1/4 mask of the simulation experiments, PSNR and SSIM are improved by 0.3214~3.6793 and 0.0005~0.0543, and average pixel L_1 loss, NIQE and BLIINDS are decreased by 0.0171~4.1120, 0.0565~4.7440 and 0.8841~4.2586, compared with other classical methods, respectively. In the real cloud-containing image experiments, NIQE and BLIINDS indexes are decreased by 0.1062~1.8992 and 1.0903~5.6495. Visual effects under the same conditions are shown and analyzed. The subjective and objective experimental results show that compared with the classical method, the proposed method has certain advantages in semantic rationality, information accuracy and visual naturalness, and provides a better solution for single spectral image pixel value prediction against thick cloud covering.

Keywords unmanned aerial vehicle image; image generation; two-stage model; deep learning; deep convolutional generative adversarial net

1 引 言

随着无人机技术的快速发展,其用途日益广泛,但拍摄过程中云层遮挡会造成图像可用信息熵下降. 通常人们按云的底部距离地面的高度不同把云分为低云(云底高度一般在 2000 m 以下)、中云(云底高度通常在 2000~6000 m 之间)和高云(云底高度通常在 6000 m 以上)三种. 本文研究的无人机含云图像拍摄于 2000 米以上高空,成像时会受到低云和中云的影响,由于较厚的云层会使被遮挡区域信息完全丧失,给大视场图像拼接及后续处理造成严重影响. 因此如何利用已有的信息最大程度地推断被云污染的图像区域信息,并使之符合图像语义统计特性和人眼视觉感知,一直是国内外学者十分关注的问题.

通常人们将图像云区内容预测视为图像复原的子问题,但厚云遮挡情况下难以无失真地重建所有细节,需要利用图像内容的相关性和先验,对遮挡区域进行生成,得到与真实像素在色彩、语义和人眼视觉上满足最大似然估计的结果[1].

与卫星、飞机、飞艇、气球等空基成像平台一样,无人机图像也是光谱、时间和空间等因素关联函数下的约束生成,但其还具有飞行高度受限、航迹随机、成像光谱单一以及航时短等特点. 针对上

述特性，考虑到目前深度学习在图像领域取得了突出的表现，本文尝试引入基于深度学习的图像修复方法并加以改进，对无人机图像厚云遮挡区域的内容进行填补生成，提出了新的模型框架和损失函数，并在无人机图像数据集上进行了对比实验验证. 本文的主要贡献有：

（1）提出了一种基于深度学习的两阶段厚云区域内容生成方法，结合图像修复和多时相法的特点，利用一阶段修复的结果，检索相似分布的图像样本，并利用注意力机制对这些图像进行对齐，避免了直接在多分布数据上修复而产生的效果不理想.

（2）提出了一种注意力对齐方法，利用 BoW 算法检索与一阶段修复结果相似分布的图像样本，设计了仿射网络对这些样本与一阶段输出结果进行空间对齐，一同进入二阶段推断网络，实现了空间上的信息互补，充分利用了图像库相似信息.

（3）改进了 DCGAN 模型的结构和损失函数，作为两阶段模型的主框架. 采用了全局—局部两个判别器对图像的生成进行约束，提出了新的联合损失函数，使生成的图像在全局和局部都具有语义合理性.

（4）相较于现有图像修复方法，在包含较大云区范围且不同地物的无人机图像上，本文方法可以更好地生成缺失部分的地物结构信息，并保持了生成部分与背景的连续性和一致性.

2　相关工作

目前针对航拍图像的云区内容估计方法可分为多光谱法、多时相法以及基于图像修复的方法.

多光谱法[2-4]多用于卫星图像. 由于云的光学特性，使得其对可见光和其它波段光谱产生不同程度的影响. 因此，可利用不同光谱间的相关性对可见光含云区域进行估计，找出光谱间的函数关系. 但选择的光谱不同，处理方法也各不相同，人们在不同成像平台，利用不同源图像估计被遮挡信息，如红外图像与可见光图像去云[2]，或合成孔径雷达图像与可见光图像去云[3]，以及对高光谱图像进行回归估计[4]等. 但多光谱法对传感器和配准算法要求很高，且对于云层较厚的情况难以处理.

多时相法[5-9]是利用不同时间段对同一地区的成像结果，对云区进行信息补全. 假设在时段集合内目标地物没有发生显著变化，且每个区域至少有一幅图像是不含云的，利用图像块匹配或直方图匹配等算法，从云区边界的纹理信息中找出这一区域的不含云图像块，进行填充和融合. 由于卫星遥感图像一般都有多光谱数据，一些方法[8-9]将多时相和多光谱结合起来，以求在更大的搜索空间中找到含云区域的更准确表示. 多时相法适合于卫星图像的精确修复，但如无人机等短航时的飞行器难以做到航线固定，因此无法应用.

基于图像修复的方法[10,11]是将云区看作图像缺失部分，采用图像其余部分的空间信息预测缺失部分. 该方法无需多光谱数据，因此更适合于无人机等单一光谱成像平台. 图像修复是指重建图像缺失或损坏的部分，使其更加完整，并保持其一致性，一般用于修复图像被破坏部分，以及移除前景物体[12]. 图像修复需要利用图像的冗余信息，寻找待修复区域与其它区域潜在的复杂映射关系，对此研究者基于各种假设提出了多种模型. 如 Bertalmio 等[12]基于有界变差假设，提出的偏微分方程的方法；Starck、Elad、Gao 等[13-15]基于重复性结构和纹理，提出的基于稀疏表示的修复方法；Criminisi、Huang 等[16-19]基于图像自相似性，提出的样本块特征匹配方法. 这些方法在缺失部分不大、结构简单、具有重复纹理的图像上表现较好，但由于所采用的特征是人工设计的，没有对图像整体分布进行学习，因此修复的效果语义合理性不足.

随着深度学习理论、卷积神经网络和生成对抗网络[20,21](Generative Adversarial Networks, GAN)的提出和发展，人们逐渐将其引入到图像修复领域. Yeh[22]在 DCGAN 的基础上提出了一种二次寻优的语义修复方法，第一步在所研究的数据集上训练一个 DCGAN 模型以学习图像语义，第二步针对图像寻找最优输入编码. 由于第二步是一个非凸优化，当第一步未能充分学习时，最终寻优的结果会严重失真，因此该方法在分布比较集中、易学的数据集上表现出色. 为避免这个问题，人们直接用含掩膜(mask)图像做输入. Pathak[23]提出基于 DCGAN 的 context encoder 模型，用 auto-encoder 结构[24]作为 G 网络生成修复图像，用 D 网络衡量 G 的输出与真实图像的差异，将欧氏距离损失和对抗损失加权作为联合损失函数. 这种直接的对抗训练简化了修复过程，G 的作用相当于回归器，而 D 相当于正则化，减弱修复图像的模糊程度. 但 Pathak[23]的模型过于简单，不足以更好地回归图像缺失部分和其它部分的函数关系，损失函数也只关注了图像整体，因此在掩膜区域产生了伪迹，需要采用新的模型进行改进.

Iizuka[25]在 Pathak[23]的基础上，增加了 G 网络的层数，D 网络增加了 1 个独立的局部分支(local D)，最后与 global D 进行全连接层级联，用来鉴别

修复区域的真伪，提升了图像局部修复效果. Yu[26]沿用了 Iizuka[25]的全局—局部结构，提出了两阶段粗—精修复模型. 该方法在一阶段粗修复后，二阶段增加了并行的注意力机制支路，利用一阶段修复结果，估计掩膜内外的相关性，因此在一些单一分布自然图像数据集上效果比较理想. 上述方法改进的是图像修复技术，研究对象是地面自然场景图像，视点、场景和需要去除的遮挡物之间距离较近，如果直接用于高视点、远场景和随机云层遮挡的无人机航拍图像修复，具有局限性.

我们曾基于 context encoder 模型针对遥感图像厚云遮挡做过类似工作[27]，由于遥感图像画幅较大，地物信息变化范围较小，因此在地物分类的基础上采用了 encoder-decoder 结构对图像的整体进行压缩和回归，并以梯度正则项控制修复的平滑性. 但由于图像本身的画幅较大，并未考虑多幅图像间信息的互补性.

上述基于修复的方法直接应用于无人机图像云区内容预测会存在以下突出问题：（1）现有的大多数修复方法仅针对 256×256 左右分辨率的单幅图像，而无人机图像是经过拼接的全景图，因此将其进行切割，由此产生了存在多种地物多种分布的数据集，现有方法对此还缺少研究；（2）无人机图像数据包含森林、田地、城镇等多种地物，而这些同类地物具有一定的相似性，利用同类相似性对图像修复具有提升的效果，而现有图像修复方法大多是针对单幅图像，没有充分利用这些信息；（3）对于大云区遮挡、多分布、多类型地物的复杂图像，修复结果往往会出现纹理断裂、预测失误、轮廓模糊，造成伪迹伪影的情况.

针对上述问题，本文结合多时相法和图像修复法的特点，提出了一种新的无人机图像厚云区域内容生成算法. 首先，将无人机全景图按照固定尺寸进行无重叠切割，得到图像数据集；然后设计了两阶段生成网络的主框架，以一阶段修复结果搜索相似样本，一同作为二阶段的输入，生成最终结果. 该结果综合了同类相似地物的特征，避免了多分布数据集学习的难题，取得了较好的效果.

3 方 法

鉴于现有深度学习的修复方法大多只能处理 256×256 左右分辨率图像，本文将无人机得到的全景图按照 256×256 的尺寸进行无重叠切割，得到图像数据集. 用 2 个 DCGAN、1 个检索模块和 1 个仿射网络相连接，设计了一个两阶段的内容生成模型，训练时，用仿真的云掩膜遮挡图像的任意区域，将原图像和掩膜图像送入一阶段修复网络进行训练，利用一阶段的输出结果，在图像数据集中检索相似的 N 个样本，再通过一个仿射网络将其与一阶段的结果图像对齐，之后将它们一同送入二阶段推断网络，在一个更大的空间对待生成区域作回归，使生成效果更加具有语义准确性和人眼视觉自然感. 测试时，先对真实含云图像进行云区检测，得到掩膜图像，直接送入 $G1$，经过前向运算得到结果，两阶段中的 $D1$、$D2$ 不再使用. 模型总体框架如图 1 所示（云区用白色标出，实箭头为运算前向传播路径，虚箭头为损失后向传播路径，下同）.

3.1 检索模块

由于地物的多样性，将无人机全景图像切割后，必然存在多个中心的分布，而要同时学习这些分布，并能根据待生成图像自动识别属于哪一种分布是十分困难的. 同时，图像内容的生成不应只借助单个地物的其它内容，还要借助其它相似地物的内容. 对此，本文采用了另一种解决方案，利用一阶段修复的结果，在数据集中检索最相似的 N 个样本，认为这些样本与一阶段输出图像具有独立同分布特性. 这 N+1 幅图像将进入二阶段推断网络中. 检索的流程如图 2 所示.

这里应用了 BoW 算法进行图像检索，事先对图像集中的所有图像进行 sift 特征提取和 K-means 聚类，形成数据集的码书，每幅图像对码书作最近距离的直方图统计，为避免码字本身频率的影响，进行词频—逆文本频率(term frequency–inverse document frequency, TF-IDF)处理，最后将直方图特征归一化作为该幅图像的特征向量. 应用时，一幅图像经过相同的特征提取，在图像集的特征上做投影和排序，选取其中最接近的样本作为检索结果.

3.2 仿射网络

大多数 CNN 模型对特征位置的平移、缩放、旋转等仿射变换不具有敏感性[28]，为了能够充分利用检索结果的语义内容，不同于 Yu[26]方法在掩膜内外的注意力机制，在检索模块后设计了仿射网络来大体上对正待修复图像和检索图像. 实验分析发现，将图像对齐后，感兴趣区域也会随之对准. 基于此建立的仿射网络将有助于更好地借助检索图像的内容进行推断[29]. 仿射网络的模型结构如图 3 所示.

以 AlexNet[30]结构为仿射参数估计网络，设 $\boldsymbol{P}$ 为待估计的仿射参数矩阵，含 6 个未知参数，对于源图像中的坐标 $(x,y)^T$，经 $\boldsymbol{P}$ 变换得到的坐标为

$(x',y')^T$，则

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \boldsymbol{P}\begin{pmatrix} x \\ y \end{pmatrix} \tag{1}$$

训练时，对每一幅源图像，人工设定任意 $\boldsymbol{P}$ 参数，对图像做(1)式变换，得到目的图像. 将源图像和目的图像输入 AlexNet 网络，得到仿射矩阵估计值 $\boldsymbol{P}'$. 由于 6 个参数之间存在尺度差异，将 $\boldsymbol{P}$ 和 $\boldsymbol{P}'$在固定的网格点采样上的变换结果之差作为损失函数

$$L_{affine} = \sum_{\boldsymbol{g}} \left\| \boldsymbol{P}\boldsymbol{g} - \boldsymbol{P}'\boldsymbol{g} \right\|_2 \tag{2}$$

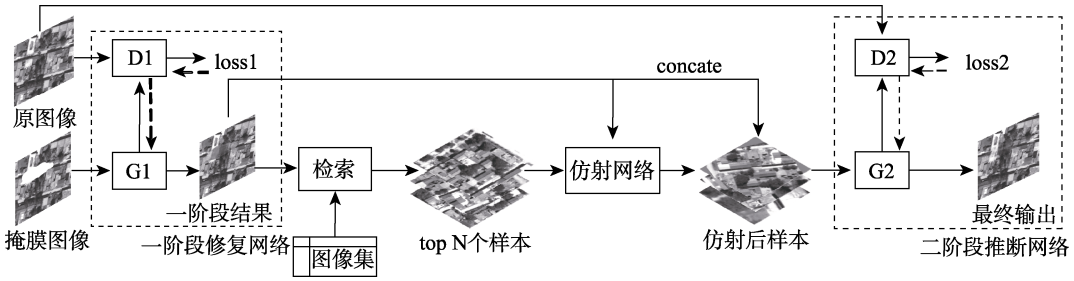
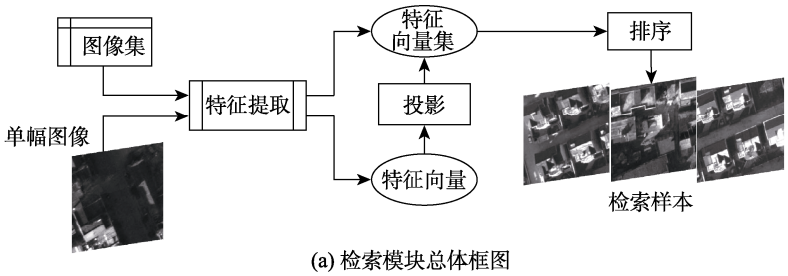
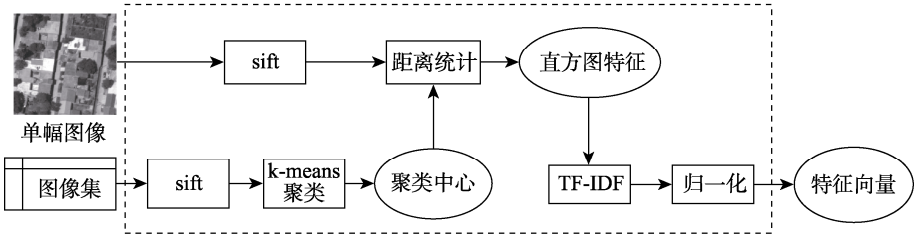


图 1　模型总体框架

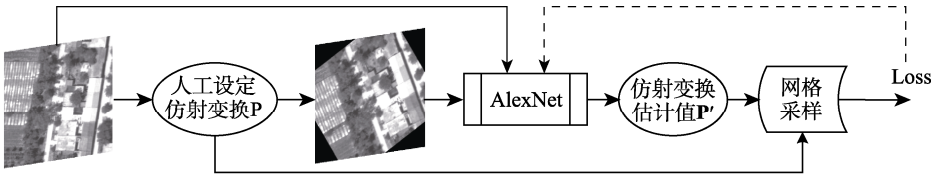


(a) 检索模块总体框图

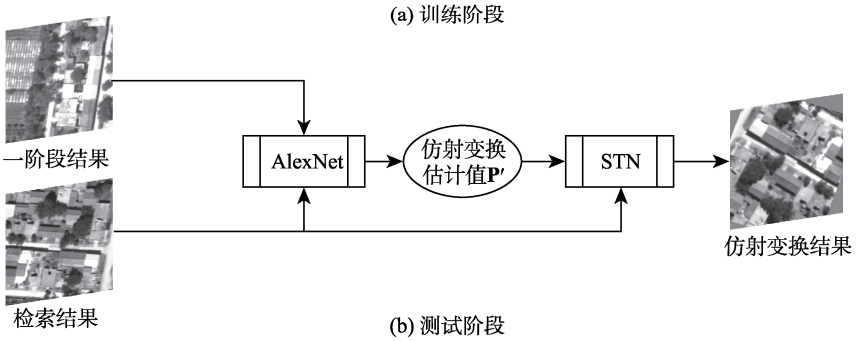


(b) 特征提取部分流程图（框中对应(a)图特征提取）

图 2　检索模块流程图



(a) 训练阶段



(b) 测试阶段

图 3　仿射网络

其中 $\boldsymbol{g}$ 为网格点坐标，这里在 256×256 的平面横纵坐标间隔 20 取点. 在训练集上不断训练 AlexNet，使之具有估计仿射参数的能力. 测试时，以一阶段结果为目的图像，以检索结果为源图像输入已训练的 AlexNet 网络，得到 $\boldsymbol{P}'$，通过空间变换网络 STN[31]，实现检索图像的仿射变换.

3.3 两阶段DCGAN网络

两阶段修复网络和推断网络的模型结构决定了图像内容生成的质量，过简单的模型参数少，回归能力不足；过复杂的模型容易在数据量不足时过拟合. 结合数据量和复杂程度等因素，经过大量验证，两阶段的模型都选用了相同的 DCGAN 结构，采用了全局—局部机制，对图像整体和生成部分的最小外接矩形区域分别用全局鉴别器(global *D*)和局部鉴别器(local *D*)进行判别，使之在全局和局部都具有语义合成能力. 具体网络结构如图 4 所示.

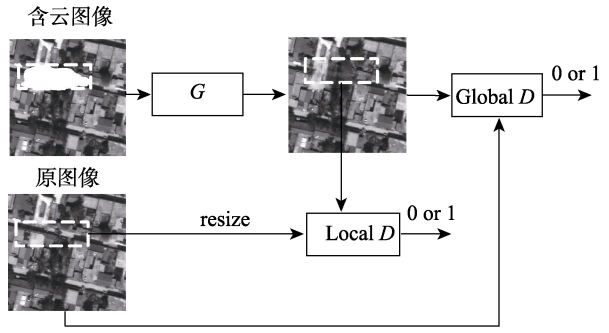


图 4 修复和推断网络相同的模型结构

不同于文献[26]方法的两阶段结构，本文一阶段模型生成初步修复结果后，将该结果输入到上述检索模块和仿射网络，得到数据集中注意力对齐的相似样本，然后将这些样本与初步修复结果进行通道维的连接后输入二阶段模型，生成最终结果. 由于无人机拍摄信息的冗余性，只要云的遮挡是局部的，在图像库中检索出的样本将大概率地包括含云图像被遮挡的信息，给第二阶段以可靠的数据支持（记为情况 1）；即使不存在该处信息，也提供了其它相似样本的特征，从而有助于图像内容的推断（记为情况 2）. 因此二阶段相当于利用其它样本的信息融合到待处理对象中.

训练时，在 256×256 的不含云原图像随机位置加上掩膜形成含云图像，输入 *G* 网络得到生成图像. Global *D* 对含云图像和原图像做整体判别，local *D* 对含云图像和原图像的云区像素做局部判别. 测试时，只需使用 *G* 网络对云区内容进行前向计算. *G* 网络、global *D* 和 local *D* 网络的结构如表 1 所示，其中形如 conv(4,2,32)表示数量为 32、大小为 4×4 的卷积核，进行步长为 2 的卷积操作，deconv(4,1/2,256)表示数量为 256、大小为 4×4 的卷积核，进行步长为 1/2 的上卷积操作，fc(3000)表示具有 3000 个节点的全连接层，bn 表示批归一化，relu、lrelu 和 tanh 分别为 3 种激活函数.

表 1 修复和推断网络结构明细

(a) *G* 网络

操作	输出
输入	256×256×1
conv(4,2,32), bn, relu	128×128×32
conv(4,2,64), bn, relu	64×64×64
conv(4,2,128), bn, relu	32×32×128
conv(4,2,256), bn, relu	16×16×256
conv(4,2,512), bn, relu	8×8×512
fc(3000), bn, relu	3000
fc(8×8×512), bn, relu	8×8×512
deconv(4,1/2,256), bn, relu	16×16×256
deconv(4,1/2,128), bn, relu	32×32×128
deconv(4,1/2,64), bn, relu	64×64×64
deconv(4,1/2,32), bn, relu	128×128×32
deconv(4,1/2,1), tanh	256×256×1

(b) global *D* 网络

操作	输出
输入	256×256×1
conv(4,2,32), bn, lrelu	128×128×32
conv(4,2,64), bn, lrelu	64×64×64
conv(4,2,128), bn, lrelu	32×32×128
conv(4,2,256), bn, lrelu	16×16×256
conv(4,2,512), bn, lrelu	8×8×512
conv(4,2,1024), bn, lrelu	4×4×1024
fc(1)	1

(c) local *D* 网络

操作	输出
输入	128×128×1
conv(4,2,64), bn, lrelu	64×64×64
conv(4,2,128), bn, lrelu	32×32×128
conv(4,2,256), bn, lrelu	16×16×256
conv(4,2,512), bn, lrelu	8×8×512
conv(4,2,1024), bn, lrelu	4×4×1024
fc(1)	1

3.4 损失函数

对于任意 GAN 模型，*G* 和 *D* 的训练是以下极大—极小联合优化问题

$$\min_G \max_D \mathrm{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log D(\boldsymbol{x})]+E_{\boldsymbol{z}\sim p_G(\boldsymbol{z})}[\log(1-D(\boldsymbol{z}))] \tag{3}$$

其中 $\boldsymbol{x}$ 为真实图像样本，$\boldsymbol{z}$ 为经过 G 生成的图像，$p_{data}(\boldsymbol{x})$ 和 $p_G(\boldsymbol{z})$ 分别为二者的概率分布函数. 对训练中某幅特定图像，有下列关系

$$\boldsymbol{z} = G[\boldsymbol{x} \odot (1-\boldsymbol{M})] \tag{4}$$

式中二值图像 $\boldsymbol{M}$ 为云区图像掩膜，只在云区为 1，$\odot$ 为对应像素相乘操作. 由于同时引入了 global D 和 local D，对 G 的联合损失函数为

$$L_G = \lambda_p l_p + \lambda_{adv\_g} l_{adv\_g} + \lambda_{adv\_l} l_{adv\_l} + \lambda_{tv} l_{tv} \tag{5}$$

式(5)右边 4 项分别为感知损失、全局对抗损失、局部对抗损失以及总变差损失[32]及其加权系数，其中感知损失定义为生成图像与原图像的 L_2 范数损失

$$l_p = \frac{1}{N}\sum_{i=1}^{N} \left\| \boldsymbol{x}_i - \boldsymbol{z}_i \right\|_2^2 \tag{6}$$

其中 N 为一批图像数量. 加入感知损失是为了在内容上接近真实图像，但仅有感知损失会导致模糊，因此还需加入对抗损失，生成接近真实语义分布的像素. 全局和局部对抗损失定义为

$$\begin{cases} l_{adv\_g} = -\log D_{global}(\boldsymbol{z}) \\ l_{adv\_l} = -\log D_{local}(\boldsymbol{z}') \end{cases} \tag{7}$$

式中 $\boldsymbol{z}'$ 表示 $\boldsymbol{z}$ 的云区部分缩放到 128×128 大小的图像. 总变差损失定义为生成图像梯度大小的平方

$$l_{tv} = \sum_{i,j} [(\boldsymbol{z}_{(i,j+1)} - \boldsymbol{z}_{(i,j)})^2 + (\boldsymbol{z}_{(i+1,j)} - \boldsymbol{z}_{(i,j)})^2] \tag{8}$$

总变差损失是正则化项，可防止生成过多噪声. 经过多次交叉验证，最后设定的加权系数为

$$\begin{cases} \lambda_p = 0.99 \\ \lambda_{adv\_g} = 0.006 \\ \lambda_{adv\_l} = 0.004 \\ \lambda_{tv} = 1\times 10^{-6} \end{cases} \tag{9}$$

对于 global D 和 local D，同样必须在训练中提高鉴别真实不含云图像和生成图像，其损失函数为

$$\begin{cases} L_{D\_global} = -\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} [\log D(\boldsymbol{x}_i) + \log(1 - D(\boldsymbol{z}_i))] \\ L_{D\_local} = -\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} [\log D(\boldsymbol{x}'_i) + \log(1 - D(\boldsymbol{z}'_i))] \end{cases} \tag{10}$$

两阶段的 DCGAN 均采用该损失函数进行训练，所不同的是二阶段网络的输入为 N+1 幅图像.

4　实　验

4.1　训练、测试及配置

实验数据采用我国中部地区无人机可见光航拍图像，原图为拼接的单通道全景图像，将其切割为大小 256×256 的图像块，经过筛选共 31600 个样本. 考虑到计算机运算资源限制，检索样本数 N 取 3. 采用人工模拟的方法创建云区掩膜，因为一般图像修复方法常采用中心面积 1/4 区域的掩膜，所以在随后也有该类型的对比实验.

将所有样本按 9:1 的比例随机划分训练集和测试集，用所有的训练集训练检索模块的词袋聚类，用一半的训练集训练仿射网络中的 AlexNet，以避免后续的训练中偏好的结果对二阶段训练不利. 训练一阶段网络时，批大小(batch size)设为 32，迭代轮数为 30 轮；二阶段 batch size 设为 16，迭代轮数为 20 轮. 训练时，随机置乱样本顺序，并随机垂直、水平反转图像以进行数据增广. 在整个训练和测试中，检索模块对每个输入图像的检索结果都去掉了原图像，以避免出现多时相法的理想条件. 学习算法采用 Adam 优化算法，实验中发现 G 和 D 网络学习速率不同，因此设置 G 更新 5 次时，D 更新 1 次，经过相应轮数的迭代，global D 和 local D 对样本的分类准确率稳定在 50%附近，认为模型的学习达到稳定.

运行主要硬件环境为 Nvidia GTX1070Ti GPU、48G 内存、4 核 xeon CPU，软件环境为 win10 系统、python3.5+tensorflow1.7.0 编程环境. 一次训练时间约为 5 天. 经对 3160 幅样本进行测试，结果表明，平均对 1 幅图像的处理时间为 0.2364s，其中一阶段修复用时 0.0566s，检索模块和仿射网络用时 0.1113s，二阶段推断用时 0.0685s.

4.2　生成结果的准确性

尽可能真实地估计被云遮挡的景象是本文方法最重要的目的. 但在实拍含云影像中得到同样视点的无云参考图像的概率很小，为了方便测试不同的方法，这里用仿真的云区掩膜覆盖测试集图像，再输入训练好的模型得到结果. 为横向对比本文方法的生成效果，选取了有代表性的基于样本块的修复方法[19]和深度学习的图像修复方法[22,23,25-27]，对同样的数据进行训练后评估. 由于文献[19]无需学习，直接将图像和二值掩膜图像送入算法得到结果. 文献[22]需要二次优化，进行 batch size 为 64 的 200 轮训练后，对每幅测试图像和二值掩膜又进行了 1000 次的迭代寻优. 对文献[23,27]方法，进行 batch size 为 32 的 100 轮训练. 文献[25]方法的 batch size 设为 64，进行了 100 轮重建损失训练、1 轮对抗损失训练和 300 轮联合损失训练. 文献[26]方法的 batch size 设为 16，1、2 阶段分别进行 100 轮训练. 上述方法的代码均来自作者主页或 github 平台. 测试

时分别采用中心掩膜和两种模拟云掩膜得到对比实验结果.

4.2.1 视觉效果对比

选取部分实验结果，分成中心掩膜和仿真云掩膜，如表 2 所示，由人眼主观判断可知本文算法取得了最优视觉效果.

由表 2 可见，文献[19]是基于样本块匹配的非学习算法，修复时是对相似块进行填充，对规则的直线方块区域修复效果好，但因为无法对数据学习，在缺乏相似块时出现明显的错误. 由于学习能力的不足，文献[22，23]方法产生了不同程度的失真；而文献[25]缺少了注意力机制，未能学习到掩膜内外的映射关系，因此修复的部分大多是像素的平均，纹理比较模糊；文献[26]的注意力机制仅限于图像本身，而由于数据分布的多样性，样本间掩膜内外的映射关系差异较大，因此出现了语义不合理的因素；文献[27]的使用条件是高空遥感图像，增加了梯度的正则，但对像素变化剧烈处难以拟合. 而本文方法的两阶段设计综合了图像本身和检索图像的互信息，因此取得了较优结果.

4.2.2 全参考指标对比

常用的评价图像生成准确性的指标为全参考图像质量评价方法，此处选择 3 种指标：PSNR、SSIM 和平均像素 L_1 损失. PSNR 与对应像素点的均方误差有关，数值越高说明与参考图像差异越小；SSIM 是图像亮度、对比度、结构三方面相似性的乘积，数值在 0~1 之间，越高说明越相似；平均像素 L_1 损失描述了对应位置像素的绝对差异，数值越小代表越好. 3 种指标反映了内容生成的信息准确性，表 3 给出了在中心掩膜下本文方法和对比方法在以上指标下的对比.

由表 3 可见，文献[25]方法在 3 种指标下评价最好，具有较好的预测准确性，但综合表 2 中结果，[25]产生了明显的模糊，与人的感知质量好坏存在一定偏差（表 4 给出了几例模糊结果指标偏高的数据），是由于全参考指标对图像模糊不敏感所致[33,34].

本文方法不仅未产生明显模糊，且在 3 种指标下取得了次优的结果. 除了文献[25]方法，本文比其它 5 种方法在 PSNR 上提高 0.3214~3.6793，在 SSIM 上提高 0.0005~0.0543，在 L_1 损失上下降 0.0171~4.1120.

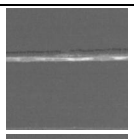
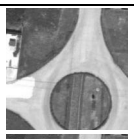
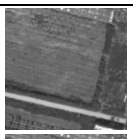
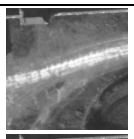
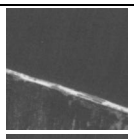
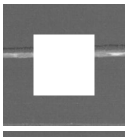
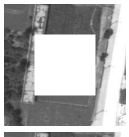
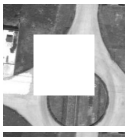
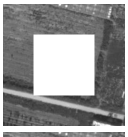
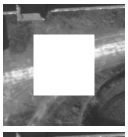
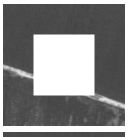
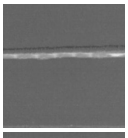
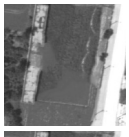
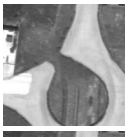
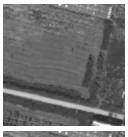
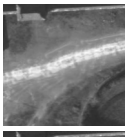
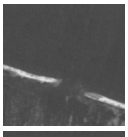
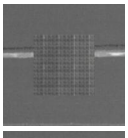
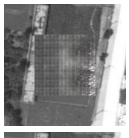
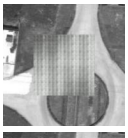
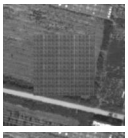
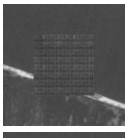
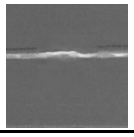
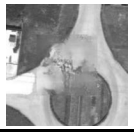
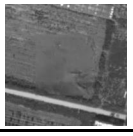
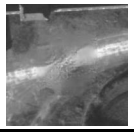
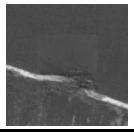
4.2.3 方法的可解释性

本文在生成准确性上的优势主要来自于检索模块和仿射网络的引入，为解释这种机制的作用，给出了 4 例测试样本的中间结果，如图 5 所示.

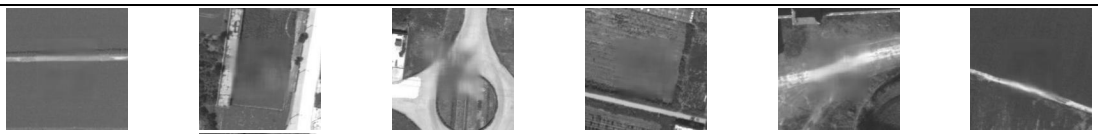
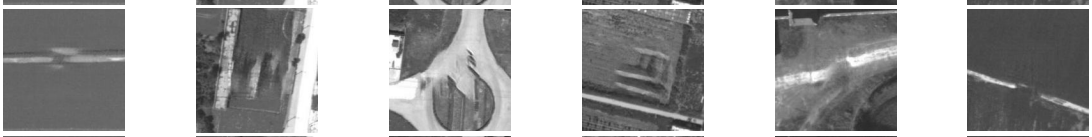
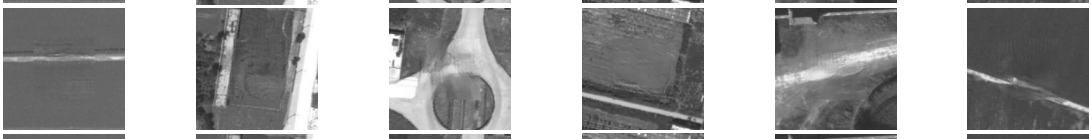
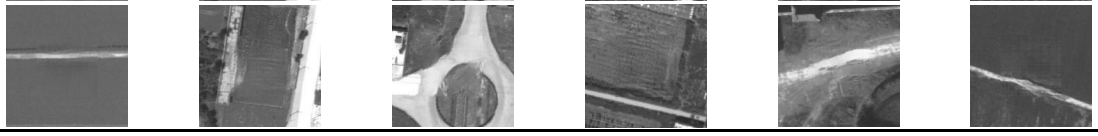
图 5(a)(b)对应于 3.3 节所述情况 2，图 5(c)(d)对应于情况 1. 图 5(a~d)第 1 行从左至右依次为原图、加云掩膜图像、一阶段和二阶段输出结果，第 2 行为检索到的 top3 图像，第 3 行为一阶段修复结果分别与 3 幅 top3 图像输入仿射网络得到的对齐结果.

表 2 同类方法主观对比

(a) 中心掩膜

方法	第 1 组	第 2 组	第 3 组	第 4 组	第 5 组	第 6 组
原图						
加掩膜图像						
[19]						
[22]						
[23]						

（续　表）

方法	第 1 组	第 2 组	第 3 组	第 4 组	第 5 组	第 6 组
[25]						
[26]						
[27]						
本文方法						

(b) 仿真云掩膜

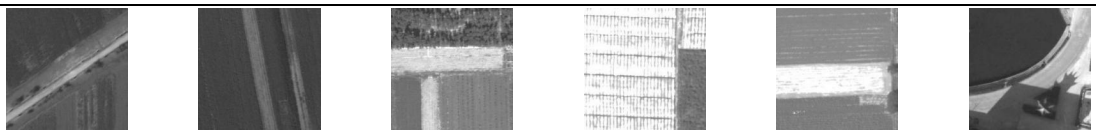
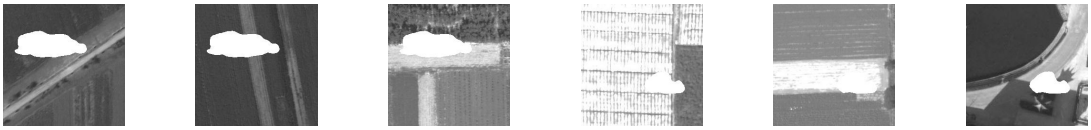
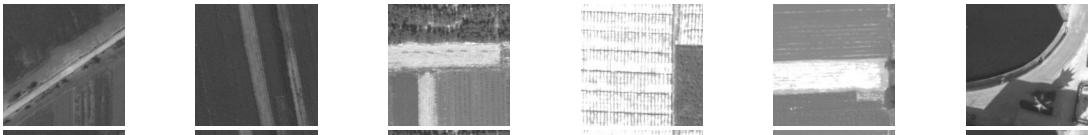
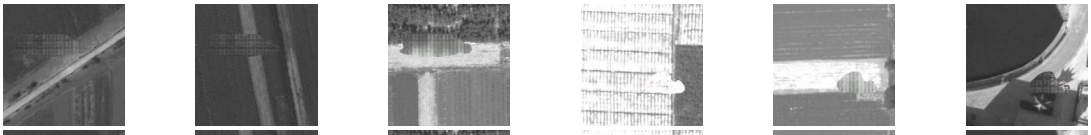
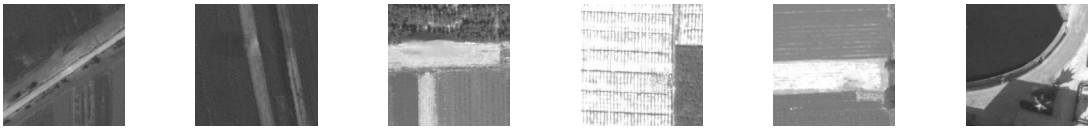
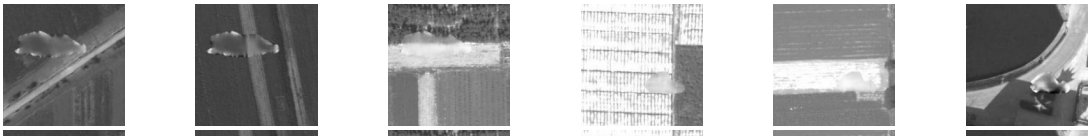
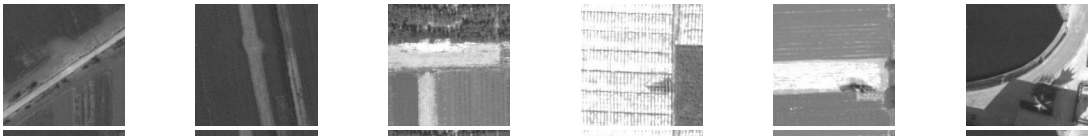
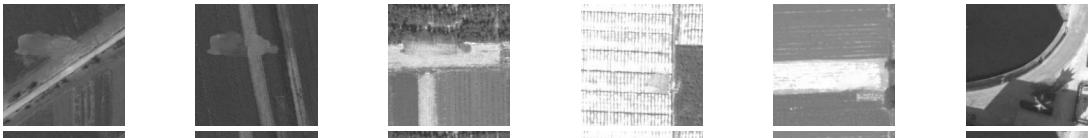
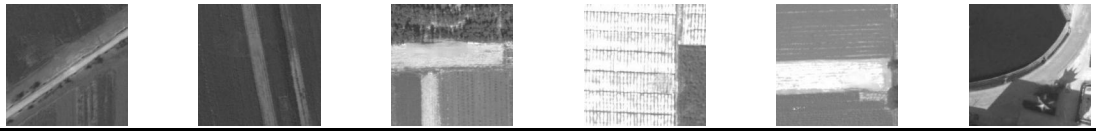
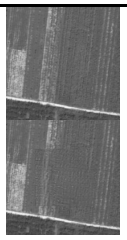
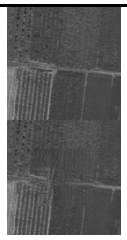
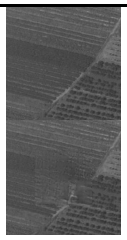
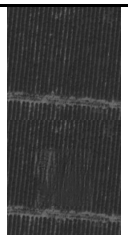
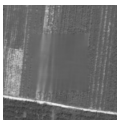
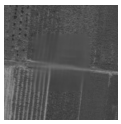
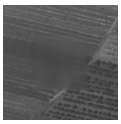
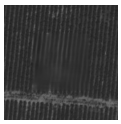
方法	第 1 组	第 2 组	第 3 组	第 1 组	第 2 组	第 3 组
原图						
加掩膜图像						
[19]						
[22]						
[23]						
[25]						
[26]						
[27]						
本文方法						

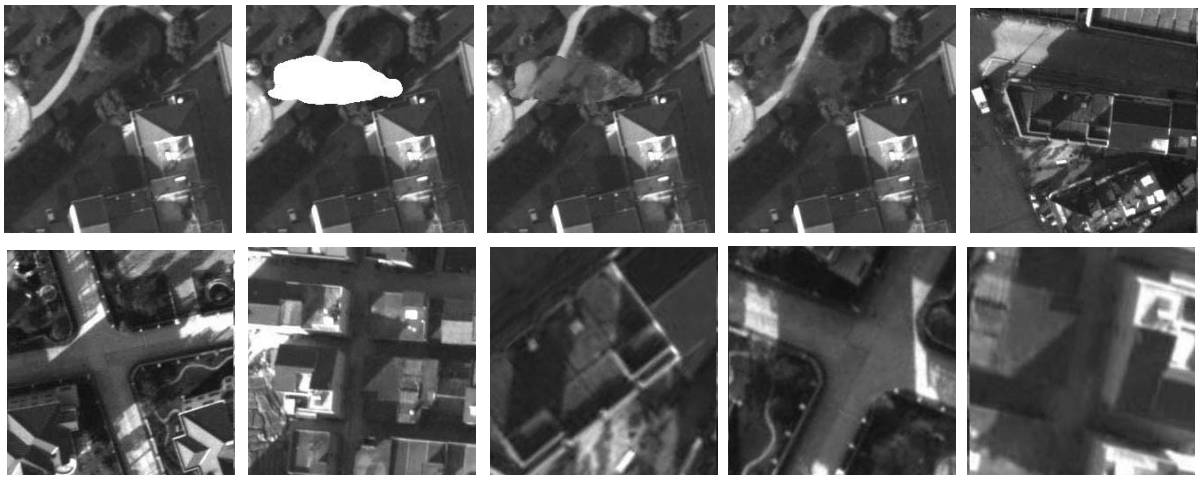
表 3 同类方法全参考质量评价对比

方法	PSNR	SSIM	L_1 损失
[19]	27.3369	0.9061	5.9271
[22]	24.1401	0.8598	8.4661
[23]	27.0765	0.9058	4.9950
[25]	28.7107	0.9357	4.1546
[26]	26.7628	0.9062	5.2512
[27]	27.4980	0.8524	4.3712
本文方法	27.8194	0.9067	4.3541

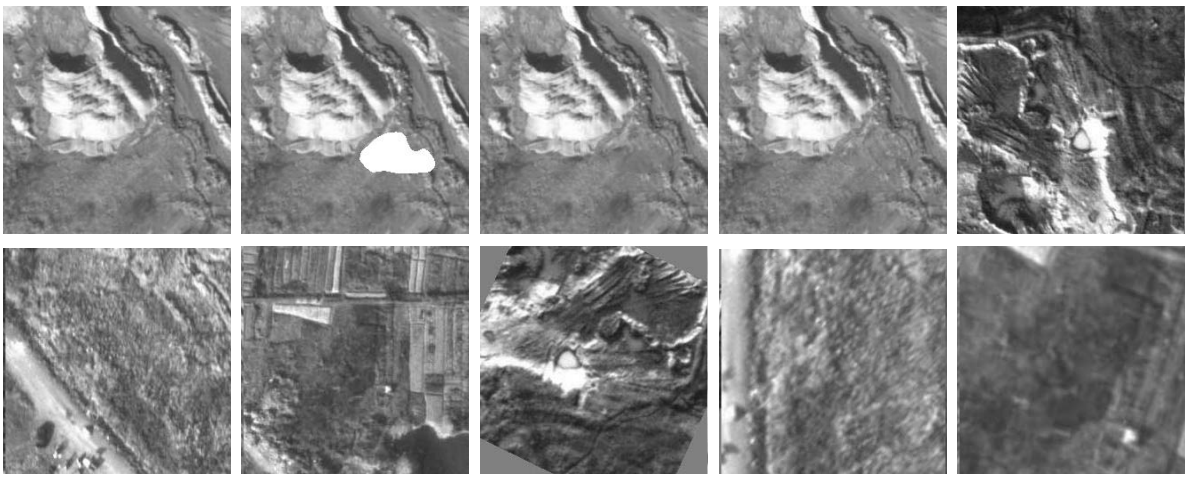
可见，仅凭单幅图像的信息，一阶段修复效果通常较差，检索得到的图像与原图像都属于同一类场景，但地物和纹理的分布和方向却有很大不同，这是因为在检索时所用的 sift 特征具有旋转、缩放和平移不变性. 经过仿射变换后，排序靠前的检索图像与原图像的相似块、纹理和结构出现的位置基本保持一致，因此二阶段推断效果在视觉和地物语义信息的完整性上均好于一阶段. 图 5(a)(b)的结果是相似分布数据的插值，图 5(c)(d)由于利用了未遮挡的图像，输出图像质量更接近原图. 在实际使用中，随着影像数量的积累，会逐渐形成类似图 5(c)(d)的信息增益.

表 4 全参考评价中模糊图像的指标值偏高现象

原图				
本文结果				
PSNR	34.2292	34.2852	35.9966	36.0403
SSIM	0.9441	0.9474	0.9599	0.9651
L_1 损失	1.8861	1.8605	1.5089	1.5194
[25]结果				
PSNR	34.3003	34.3354	36.1599	37.2439
SSIM	0.9537	0.9568	0.9654	0.9722
L_1 损失	1.8121	1.8019	1.4388	1.3324

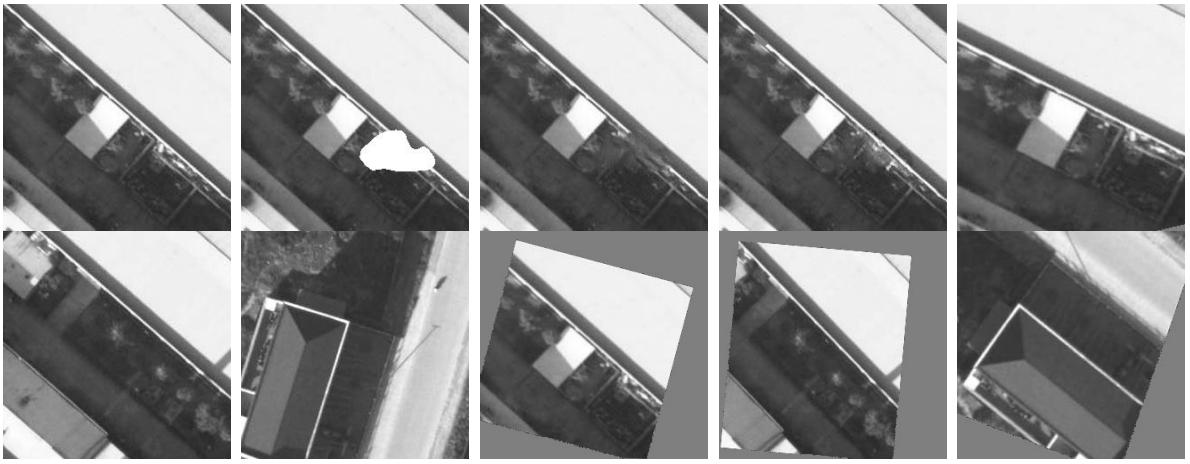


(a) 第 1 组



(b) 第 2 组

(c) 第 3 组



(d) 第 4 组

图 5　4 幅图像处理的中间结果

4.3　生成结果的自然性

云区内容生成的好坏不仅与语义合理性和准确性有关，还要兼顾人的主观感知，避免噪声、畸变和模糊失真，而且在实际中，没有参考图像作对比，因此有必要对填补结果进行无参考质量评价. NIQE[35]和 BLIINDS[36]是常用的评价方法. 它们属于自然场景统计(Natural Scene Statistics, NSS)方法，利用 2 个尺度上分块拟合的广义高斯分布系数作为特征. NIQE 将图像空间块归一化，计算特征向量与无失真图像特征间的多元高斯分布距离度量失真程度，而 BLIINDS 则先进行 DCT 变换，再将高层次特征与平均主观得分差异(DMOS)进行贝叶斯回归. 这 2 种指标的输出都表示图像的失真程度（数值越小代表质量越好）. 它们在图像处理领域广泛使用，且分别在空域和变换域处理，具有一定的互补性，本文采用它们评估内容生成的质量，平均得分如表 5 所示.

表 5 显示本文方法在 NIQE 上好于其它方法 0.0565~4.7440，在 BLIINDS 上好于其它方法 0.8841~4.2586. 综合本节与 4.2 节结果,说明本文方法兼顾了视觉自然性和语义、信息准确性，在主客观评价上都取得了最佳表现.

表 5　各方法与真实值的无参考质量评价对比

方法	NIQE	BLIINDS
Ground truth	4.1112	14.5534
[19]	6.2328	16.5456
[22]	9.2358	19.9201
[23]	4.9841	19.7373
[25]	5.7747	19.1464
[26]	5.7882	17.5071
[27]	4.5483	18.5347
本文方法	4.4918	15.6615

4.4　真实含云图像实验

为进一步验证本文方法在真实含云无人机图像的生成效果，在数据集中选取了厚云覆盖的场景，

用云检测算法[37]分割出云区作为掩膜，送入已训练好的模型. 这里选取 5 例不同高度、不同地物场景图像，原图和检测出的云区掩膜如图 6 所示，各种对比方法的处理结果如表 6 所示，结果图像的无参考评价指标平均值如表 7 所示.

从表 6 实验结果分析发现，由于厚云图像同时存在薄云和薄雾，故文献[23]方法在修复时会引入相应位置的噪声，在修复厚云区域时出现了不同程度的模糊和失真；文献[22]由于未能学习图像的分布，故生成杂乱的纹理；文献[19]虽然修复部分比较自然，但仅仅复制了原图其它部分的低层特征，因此破坏了图像的整体语义. 本文方法统一考虑了待生成区域与整体结构的关系，合理运用了两阶段模型，对图像的语义保留得比较完整，且在视觉上取得了较好的效果，同时表 7 的无参考评价的实验结果也验证了本文方法的有效性，在 NIQE 和 BLIINDS 指标上分别改善了 0.1062~1.8992 和 1.0903~5.6495，这与仿真云掩膜实验结论相一致.

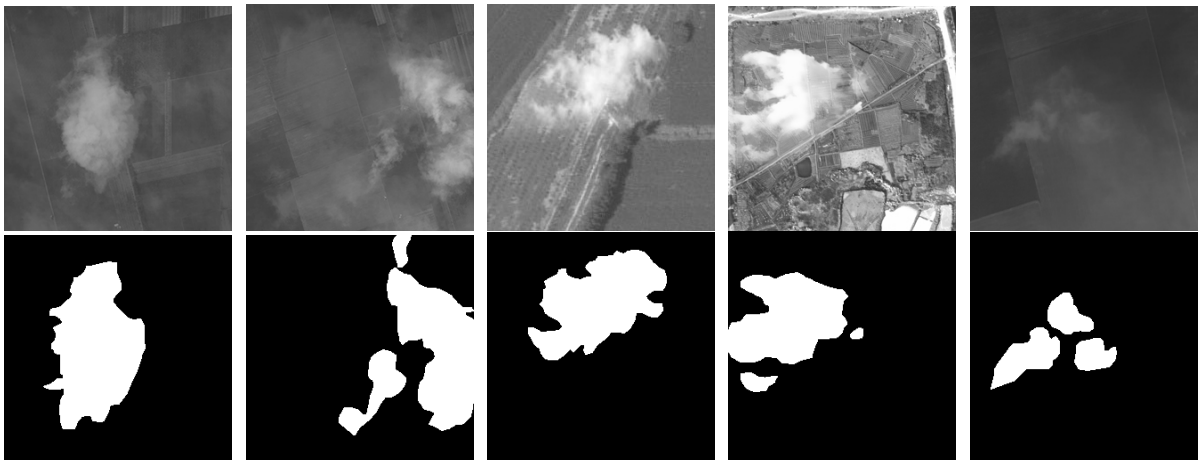


图 6 含云原图及云区掩膜

表 6 各方法对图 6 处理效果对比

方法	第 1 组	第 2 组	第 3 组	第 4 组	第 5 组
[19]					
[22]					
[23]					
[25]					
[26]					
[27]					
本文方法					

4.5 仿射网络对比实验

为验证仿射网络的必要性，保持原模型其它部分不变，仅去掉仿射网络，将检索得到的相似样本图像直接与一阶段修复图像在通道维连接后送入二阶段. 保持损失函数不变，重新训练二阶段网络，得到的部分结果与原模型对比如表 8 所示，这里仍取 N=3.

表 7 真实含云图像无参考评价平均值

方法	NIQE	BLIINDS
[19]	6.0457	15.1341
[22]	6.6879	18.4502
[23]	5.9076	15.1015
[25]	5.3691	18.1569
[26]	4.8949	13.8910
[27]	5.4674	14.7311
本文方法	4.7887	12.8007

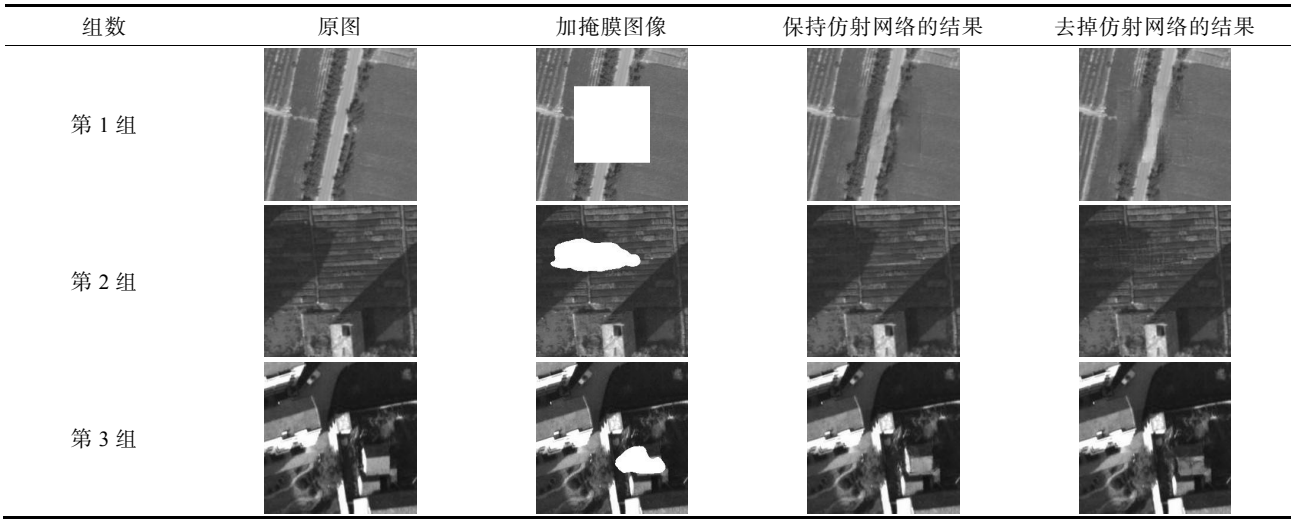
表 8 选取了 3 组不同掩膜的结果，(a)为去掉仿射网络前后的效果对比，(b)为去掉仿射网络前后进入二阶段网络的检索样本. 仿射网络将检索到的样本旋转、缩放、平移到最匹配的位置，如第 1 组道路、第 2 组田野的纹理及第 3 组房屋的结构，被旋转到相近的角度. 根据地物的相似性，推断网络能够在通道维上利用检索样本的信息，生成一致性和

连续性更好的结果.

表 9 给出了不加仿射网络时，对所有测试样本的全参考和无参考的评价结果的平均得分，相对于有仿射网络，在几种评价指标有不同程度下降.

表 8　去掉仿射网络前后的效果对比

(a) 生成效果对比



(b) 检索图像经过仿射网络前后对比

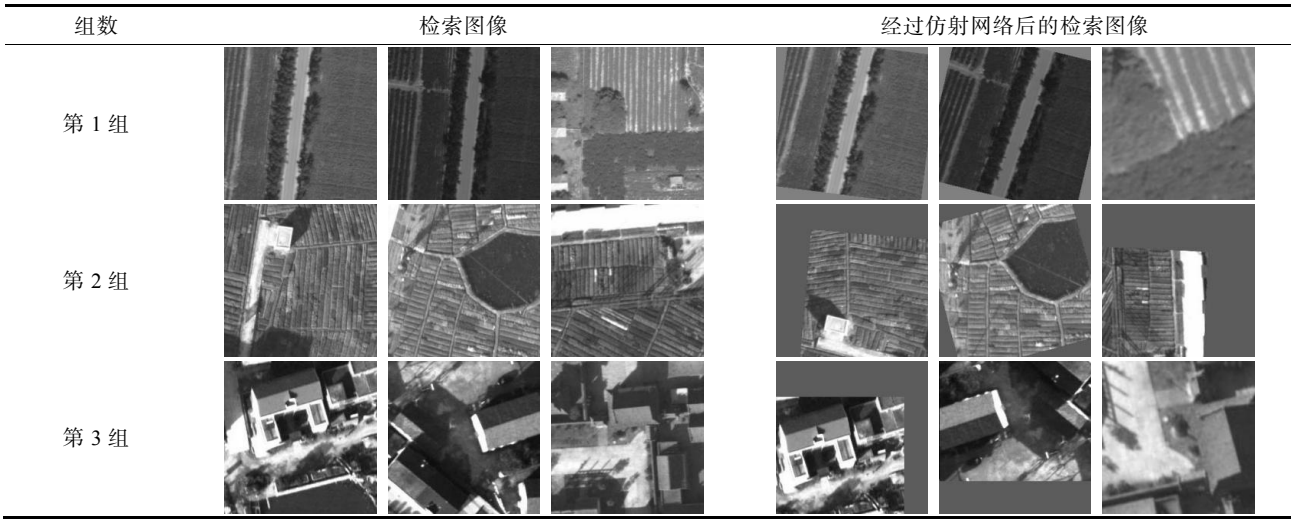


表 9　去掉仿射网络全参考和无参考质量评价结果

指标	PSNR	SSIM	L_1 损失	NIQE	BLIINDS
数值	26.2752	0.8753	4.5233	4.7188	15.7917
性能下降	1.5442	0.0314	0.1692	0.2270	0.1302

5　结　论

本文基于 DCGAN 框架，提出了一种基于两阶段模型的无人机图像厚云遮挡区域内容生成方法. 仿真云掩膜和真实含云图像实验表明，利用一阶段提供的指导信息及注意力对齐机制，使得二阶段具备了较强的语义推断能力，可较好地生成大云区遮挡、多分布、多类型地物的无人机含云图像，为更充分地利用含云影像信息提供了一种较好的选择. 与其它方法相比，本文方法具有较强的信息准确性和语义真实性，且更符合人眼主观感受.

目前本文在设计网络模型时，固定了进入二阶段网络的检索样本数量，灵活性有限，下一步将尝试根据检索到样本与待处理图像的相似度大小自适应选择样本数目，并进行不同层次的特征提取和融合，再送入相对固定的网络处理通道. 另外无人机图像还会出现视场大部分或全部被厚云覆盖的情况，这是本文目前尚未解决，后续需要进一步研究的问题.

致　谢　感谢审稿专家对本文提出的宝贵意见和建议，帮助作者对本文进行提炼与完善!

参 考 文 献

[1] Yue Z, Yong H, Zhao Q, et al. Variational denoising network: toward blind noise modeling and removal//Advances in Neural Information Processing Systems, Vancouver, Canada, 2019: 1690-1701

[2] Wang Z, Jin J, Liang J, et al. A new cloud removal algorithm for multi-spectral images//Proceedings of SPIE - The International Society for Optical Engineering, 2005, 6043: 60430W-60430W-11

[3] Huang B, Li Y, Han X, et al. Cloud removal from optical satellite imagery with SAR imagery using sparse representation. IEEE Geoscience & Remote Sensing Letters, 2017, 12(5): 1046-1050

[4] Zhang C, Li W, Travis D J. Restoration of clouded pixels in multispectral remotely sensed imagery with cokriging. International Journal of Remote Sensing, 2009, 30(9): 2173-2195

[5] Jiao Q, Luo W, Liu X, et al. Information reconstruction in the cloud removing area based on multi-temporal CHRIS images// Proceedings of SPIE–The International Society for Optical Engineering, 2007, 6790: 679029

[6] Tseng D C, Tseng H T, Chien C L. Automatic cloud removal from multi-temporal SPOT images. Applied Mathematics and Computation, 2008, 205(2): 584-600

[7] Lin C H, Tsai P H, Lai K H, et al. Cloud Removal From Multitemporal Satellite Images Using Information Cloning. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51(1): 232-241

[8] Melgani, F. Contextual reconstruction of cloud-contaminated multitemporal multispectral images. IEEE Transactions on Geoscience and Remote Sensing, 2006, 44(2): 442-455

[9] Shen H, Wu J, Cheng Q, et al. A spatiotemporal fusion based cloud removal method for remote sensing images with land cover changes. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(3): 862-874

[10] Siravenha A C, Sousa D, Bispo A, et al. Evaluating inpainting methods to the satellite images clouds and shadows removing// International Conference on Signal Processing, Image Processing, and Pattern Recognition. Jeju Island, Korea, 2011: 56-65

[11] Lorenzi L, Melgani F, Mercier G. Inpainting Strategies for Reconstruction of Missing Data in VHR Images. IEEE Geoscience & Remote Sensing Letters, 2011, 8(5): 914-918

[12] Bertalmio M, Sapiro G, Caselles V, et al. Image inpainting //Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. New Orleans, USA, 2000: 417-424

[13] Starck J L, Elad M, Donoho D L. Image decomposition via the combination of sparse representations and a variational approach. IEEE Transactions on Image Processing, 2005, 14(10): 1570-1582

[14] Mairal J, Elad M, Sapiro G. Sparse representation for color image restoration. IEEE Transactions on Image Processing, 2007, 17(1): 53-69

[15] GAO Cheng-Ying, XU Xian-Er, LUO Yan-Mei, et al. Object image inpainting based on sparse representation. Chinese Journal of Computers, 2019, 42(9): 1953-1965 (in Chinese)
(高成英, 徐仙儿, 罗燕媚等. 基于稀疏表示的物体图像修复. 计算机学报, 2019, 42(9): 1953-1965)

[16] Criminisi A, Pérez P, Toyama K. Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212

[17] Huang J B, Kang S B, Ahuja N, et al. Image completion using planar structure guidance. ACM Transactions on Graphics, 2014, 33(4): 1-10

[18] Qiang Z, He L, Xu D. Exemplar-based pixel by pixel inpainting based on patch shift//Proceedings of the CCF Chinese Conference on Computer Vision. Tianjin, China, 2017: 370-382

[19] Darabi S, Shechtman E, Barnes C, et al. Image melding: combining inconsistent images using patch-based synthesis. ACM Transactions on Graphics, 2012, 31(4): 1-10

[20] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2014: 2672- 2680

[21] Radford A, Metz L, Chintala S, et al. Unsupervised representation learning with deep convolutional generative adversarial networks// Proceedings of the International Conference on Learning Representations. San Juan, Puerto Rico, 2016

[22] Yeh R A, Chen C, Yian Lim T, et al. Semantic image inpainting with deep generative models//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 5485-5493

[23] Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: Feature learning by inpainting//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 2536-2544

[24] Masci J, Meier U, Cireşan D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction//Proceedings of the International Conference on Artificial Neural Networks. Espoo, Finland, 2011: 52-59

[25] Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. ACM Transactions on Graphics, 2017, 36(4): 107-120

[26] Yu J, Lin Z, Yang J, et al. Generative image inpainting with contextual attention//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 5505-5514

[27] LI Cong-Li, ZHANG Si-Yu, WEI Zhe, et al. Thick Cloud removal for aerial images based on deep convolutional generative adversarial networks. Acta Armamentarii, 2019, 40(7): 1434-1442 (in Chinese)
(李从利, 张思雨, 韦哲等. 基于深度卷积生成对抗网络的航拍图像去厚云方法. 兵工学报, 2019, 40(7): 1434-1442)

[28] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 3856-3866

[29] Zhao Y, Price B, Cohen S, et al. Guided image inpainting: Replacing an image region by pulling content from another image//Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision. Hawaii, USA, 2019: 1514-1523

[30] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks//Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA, 2012: 1097-1105

[31] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer

networks// Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2015: 2017-2025

[32] Yang C, Lu X, Lin Z, et al. High-resolution image inpainting using multi-scale neural patch synthesis//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 6721-6729

[33] Blau Y, Mechrez R, Timofte R, et al. The 2018 PIRM challenge on perceptual image super-resolution//Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018: 334-355

[34] Galteri L, Seidenari L, Bertini M, et al. Deep generative adversarial compression artifact removal//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 4826-4835

[35] Mittal A, Soundararajan R, Bovik A C. Making a “completely blind” image quality analyzer. IEEE Signal Processing Letters, 2012, 20(3): 209-212

[36] Saad M A, Bovik A C, Charrier C. Blind image quality assessment: A natural scene statistics approach in the DCT domain. IEEE Transactions on Image Processing, 2012, 21(8): 3339-3352

[37] Zhang Q, Xiao C. Cloud detection of RGB color aerial photographs by progressive refinement scheme. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(11): 7264-7275

WEI Zhe, master, lecturer. His research interests include deep learning, image inpainting and quality evaluation.

LI Cong-Li, master, professor. His research interest focuses on intelligent image information processing.

SHEN Yan-An, Ph. D., professor. His research interest focuses on unmanned aerial vehicle system engineering.

LIU Yong-Feng, Ph. D., associate professor. His research interest focuses on remote sensing image processing.

ZHOU Pu-Cheng, Ph. D., associate professor. His research interest focuses on image processing and analysis.

Background

Thick cloud region content prediction at UAV image is an important research direction in aerial image analysis, which can help people get cloudless panoramic image for geological exploration, disaster prevention and military reconnaissance using existing information.

Existing aerial image cloud region estimating method could be divided into three categories: multi-spectral methods, multi-temporal methods and methods based on image inpainting. Multi-spectral methods try to find the cross-correlation between every spectral imaging for several optical bands can penetrate clouds. However, multi-spectral sensor is too expensive to deploy on UAV. Multi-temporal methods search and match images from same area at different time, and compensate cloudless regions from each other. However, a fixed flight path is needed and not suitable for UAV. Methods based on image inpainting regard the covered regions as the missing parts of the image and fill in them only by spatial information. Inpainting process is a multiple regression problem. There are two main problems in current inpainting methods: unable to learn multiple distributions effectively and barely to use information from other similar samples, which restrict the semantic authenticity and pixel accuracy of the generated image.

Due to the problems, with the help of the inpainting method, we design a 2-stage end-to-end model to learn from thousands UAV images based on assumption that homogeneous UAV images have similar features at missing part. By this model, DCGAN can learn from not only the rest of the image but also the information from homogeneous samples. We also improve the structure of DCGAN as well as the loss function so that it has strong semantic synthesis ability both globally and locally. In experiments both on real and simulated data, our method performs well on NIQE, BLIINDS, PSNR, SSIM, L_1 loss and subjective feeling, proving the advantages in semantic rationality, information accuracy and visual naturalness.

This work comes from a key military project, which is dedicated to UAV reconnaissance image enhancement and target detection under adverse weather conditions, and is also supported by the Anhui Provincial Natural Science Foundation (No. 1908085MF208). Our research group has published 5 papers, 1 monograph, and declared 2 national patents related to this direction. The results of this paper are mainly to solve the problem of UAV image information enhancement under cloudy weather conditions.