文章编号　1004-924X(2025)16-2630-19

利用伪图像生成的精准侧扫描声呐图像分类

徐　涛[1*]，王朋帅[2]，郭文涛[3]，蔡　磊[1]，郑建锋[1]，陈亚晨[3]
(1. 河南科技学院 人工智能学院，河南 新乡 453003；
2. 河南科技学院 机电学院，河南 新乡 453003；
3. 中国兵器工业标准化研究所 兵工标准化研究室，北京 100089)

摘要：针对现有侧扫描声呐图像分类网络因高质量侧扫描声呐训练样本严重不足，导致对独立于训练样本以外的特定目标分类任务效果不佳的问题。本文提出一种基于特征削弱和图像块的高质量伪侧扫描声呐图像生成方法。首先，构建针对光学内容图像颜色、纹理等与声学图像非关联信息的特征削弱模块，通过淡化光学内容图像特定领域的特征，实现伪图像目标特征与声学图像特征更加相似的目的；其次，提出一种基于图像块的风格迁移伪图像生成方法，实现光学内容特征与声学风格特征的有效融合，生成兼具光学内容特征的原始伪侧扫描声呐图像；最终，引入快速引导滤波模块，对原始伪侧扫描声呐图像做增强引导，生成更接近真实水声成像环境的高质量伪侧扫描声呐图像。然后，利用大量生成的高质量伪图像扩充分类网络的训练样本，提升对真实水下环境中侧扫描声呐图像分类任务的准确率。对比实验结果表明，本文方法各项风格化评价指标平均值均表现最优，相较次优方法，风格损失、LPIPS、内容损失、SSIM 四项指标的平均值分别获得了 31.60%，3.52%，15.18%，17.55% 的性能增益。在灰度 KLSG 数据集上，分类任务的全局精度和平均精度分别达到 86.35% 和 78.54%。

关　键　词：侧扫描声呐图像；特征削弱；风格迁移；图像生成；零样本图像分类
中图分类号：TP394.1；TH691.9　　**文献标识码**：A
doi：10.37188/OPE.20253316.2630　　**CSTR**：32169.14.OPE.20253316.2630

Accurate side-scan sonar image classification using pseudo-image generation

XU Tao[1*], WANG Pengshuai[2], GUO Wentao[3], CAI Lei[1], ZHENG Jianfeng[1], CHEN Yachen[3]

(1. *School of Artificial Intelligence, Henan Institute of Science and Technology, Xinxiang* 453003, *China*;
2. *School of Mechanical and Electrical Engineering, Henan Institute of Science and Technology, Xinxiang* 453003, *China*;
3. *China Ordnance Industrial Standardization Research Institute, Ordnance Industry Standardization Laboratory, Beijing* 100089, *China*)
* *Corresponding author, E-mail: xutao@hist.edu.cn*

Abstract: The existing side-scan sonar (SSS) image classification network faced a serious issue. It suf-

收稿日期：2025-04-25；**修订日期**：2025-07-16.
基金项目：国家自然科学基金资助项目(No. 62273132)；河南省重点研发专项资助项目(No. 231111220700, No. 241111110200)；河南省高校科技创新团队支持计划资助项目(No. 25IRTSTHN018)；河南省中原创新领军人才计划资助 项目(No. 254000510043)

fered from a shortage of high-quality SSS training samples. This led to poor performance in classifying specific targets outside the training data. A high-quality pseudo side-scan sonar image generation method was proposed based on feature weakening and image patches. Firstly, a feature weakening module was designed to suppress irrelevant optical content features, such as color and texture. By weakening specific domain features in optical images, it helped adjust pseudo-image target features. This made them resemble acoustic image features more closely. Secondly, an image patch-based style transfer pseudo-image generation method was proposed. This method achieved effective fusion of optical content and acoustic style features. It generated original pseudo-SSS images with optical content features. Finally, a fast guided filtering module was introduced. It enhanced and guided the original pseudo SSS images, which generated high-quality pseudo-SSS images closer to the real underwater acoustic imaging environment. Then, a large number of high-quality pseudo images were generated. These images expanded the training samples of the classification network. This improved the accuracy of SSS image classification tasks in real underwater environments. The results of comparative experiments are presented. The proposed method achieves the best average values across various stylization evaluation indicators. Compared with suboptimal methods, the proposed method achieved significant performance gains. The average values of style loss, LPIPS, content loss, and SSIM improved by 31.60%, 3.52%, 15.18%, and 17.55%, respectively. On the grayscale KLSG dataset, the classification task achieved a global accuracy of 86.35%. The average accuracy reached 78.54%.

Key words: side-scan sonar image; feature weakening; style transfer; image generation; zero-shot image classification

1　引　言

近年来,基于侧扫描声呐(Side-Scan Sonar, SSS)图像的水下探测技术在多个领域得到了推广和应用并取得了出色的成果[1]。然而,水下声呐系统因其所用设备的特殊性,在作业时需配备经验丰富的专业技术操作员,且船只和装备必须经过特殊筛选,这导致数据采集的成本高昂且耗时[2]。一些水下声呐应用涉及国家安防等敏感信息[3],数据采集过程受到了严格管制,使得可用的数据集相对稀缺。虽然深度学习在SSS图像分类方面取得了值得称道的成果,但获取此类数据的挑战以及现有数据集的有限可用性仍然是紧迫的问题[4]。由于水下环境复杂存在各种噪声,使得侧扫描声呐图像不仅具有高噪声、低分辨率等[5]特点,且与水下光学图像一样具有失真严重、细节模糊[6-7]等问题,这对提取目标的显著特征造成困难,不利于目标分类。

受用于光学图像的样本增强方法[8]的启发,一些学者探索了水下目标识别中SSS目标图像的样本增强方法。Ma等[9]采用裁剪、旋转、调整对比度等传统方法进行数据集扩充,并以此生成对抗性示例以欺骗分类器,进而提高分类精度。Gao等[10]通过结合随机旋转、添加噪声、翻转和亮度调整来生成新图像,以满足深度学习对数据量以及海底环境的多样性的要求。Xie等[11]采用先进的分类模型,虽实现了较高的SSS声呐目标分类精度,但采用了基础的数据增强方法,未能有效解决因样本数据不足、多样性差而造成的性能限制。因此,仅使用基础的图像变换方法来扩充样本数据集,会造成实验样本缺乏足够的代表性,尽管实现了较高的分类精度,但限制了模型的泛化能力,无法对测试数据以外的样本类别进行分类。

由于海底目标的多样性和复杂性,SSS图像数据的采集操作繁琐,耗时且成本高[12]。水下目标识别迫切需要一种能够对没有预先存在的样本可供模型训练的SSS图像分类方案[13]。因此,图像生成方法成为重要的解决途径。Li等[14]提出一种基于ZSL-SSS风格迁移网络,在风格图像的指导下利用光学图像生成伪声呐图像。由于该方法存在不稳定特性,导致伪图像存在失真现

象，降低了分类效果。Li等[15]提出一种无需预训练的图像特征迁移方法，通过白化和着色操作匹配内容与风格特征，提升风格化图像的迁移效果，但生成图像存在渲染不充分的问题。Li等[16]提出一种照片级风格迁移方法，通过风格化和平滑两个步骤，能够实现无伪影的真实感风格迁移。Yao等[17]通过自注意力机制和多尺度风格交换协调空间注意分布，实现了细节层次丰富的风格化效果。Deng等[18]通过特征对齐和自适应融合实现风格迁移，在保持时间连贯性的同时精准传递艺术风格，但生成的风格化图像存在不同程度的失真。An等[19]通过可逆神经流和无偏特征传输模块防止风格迁移中的内容泄漏，确保在风格化过程不破坏原始图像内容结构。Chen等[20]通过内部统计与外部数据集学习相结合和引入双重对比损失优化风格化网络，有效解决了生成图像的伪影问题。Huang等[21]提出一种实时任意风格迁移方法，通过自适应实例归一化对齐内容与风格特征统计量，能够支持任意风格的转移，为声呐数据的扩充提供了理论支持。Natarajan等[22]结合扩散模型与大型语言模型的合成声呐框架，通过文本引导的多模态学习机制实现了多样化声呐图像的生成。Li等[23]提出两步迁移方法，先获取光学渲染图像，再通过改进UA-CycleGAN合成高质量声呐数据。Bai等[24]提出了一种能够产生与缺失类别相对应的伪SSS图像的全局上下文外部注意力网络，该方法实现了较高的分类准确率，但其伪图像存在目标纹理特征过于突出的现象。虽然这些基于注意力机制的图像生成模型对SSS图像分类效果有所增益，仅是通过扩充训练样本的数量来提高分类识别训练精度，并未真正考虑到伪样本与真实SSS图像在目标纹理、颜色、轮廓等结构特征上的相关性，且还具有不合理的局部输出和伪像[25]。本方法能参考真实SSS图像对内容图像的目标纹理、颜色等特征进行削弱，使生成的伪图像与真实SSS图像更加相似。

根据对SSS图像与光学图像的研究，虽然二者成像原理和艺术风格不同，但两种图像在部分区域的特征有相同之处。因此，任何可见的特定类别，都可以光学图像为内容数据集，任意风格的真实SSS图像为风格数据集作为输入来生成伪SSS图像。将生成图像作为训练样本来训练深度神经网络，将零样本学习问题转换成传统的监督学习问题[24]。

为了生成高质量的伪图像，实现精准的SSS图像分类，本文提出一种伪SSS图像生成网络方法，该方法包括内容图像特定领域特征削弱模块（Specific Domain Feature Weakening Module for Content Images，SDFWMCI）、基于图像块的风格迁移伪图像生成方法（Image Patch-based Style Transfer Pseudo-image Generation Method，IPSTPG）、快速引导滤波增强模块（Fast Guided Filtering Enhancement Module，FGFE）。该方法能有效缓解SSS图像分类样本不足的困境，实现高精度目标分类。

2　利用伪图像生成的SSS图像分类

根据光学样本增强研究和风格迁移的设计初衷，本研究认为生成伪图像可有效解决SSS图像样本不足的难题。本文设计了一种基于伪图像生成的SSS图像分类方法。如图1所示，本文的伪图像生成模型与AdaIN[21]和MCC[18]模型生成图像的对比图，本文方法的伪图像目标不具有目标本身的颜色和详细结构纹理的干扰，同时模拟了水下噪声阴影的分布，这与真实SSS图像的风格相似。AdaIN直接在全局范围内调整内容特征的二阶统计量，但内容细节丢失，导致风格化图像混乱。MCC采用自关注变换公式提取信道相关特征，但非线性操作的缺乏限制了网络输出的最大值，导致不适当的颜色和阴影，与声学成像的特征不相符。当生成的伪图像与真实

图1　伪图像生成结果及对比

Fig. 1　Pseudo image generation results and comparison

SSS图像的差异越小，可利用从伪图像中学习到的知识来精确地进行SSS图像分类的效果越好。

本文提出的利用伪图像生成的精准侧扫描声呐图像分类方法，算法实现分为4个步骤：(1)获取特征削弱后的光学内容图像；(2)将光学图像和SSS风格图像分割成图像块，并经过各自的编码器处理后，运用多层解码器对内容图像块进行声呐图像风格化处理。最后使用CNN解码器来获得初始伪SSS图像；(3)每个初始伪图像通过引导增强生成高质量伪图像；(4)利用伪图像训练分类模型，并使用真实SSS图像进行分类验证。总体框架如图2所示。

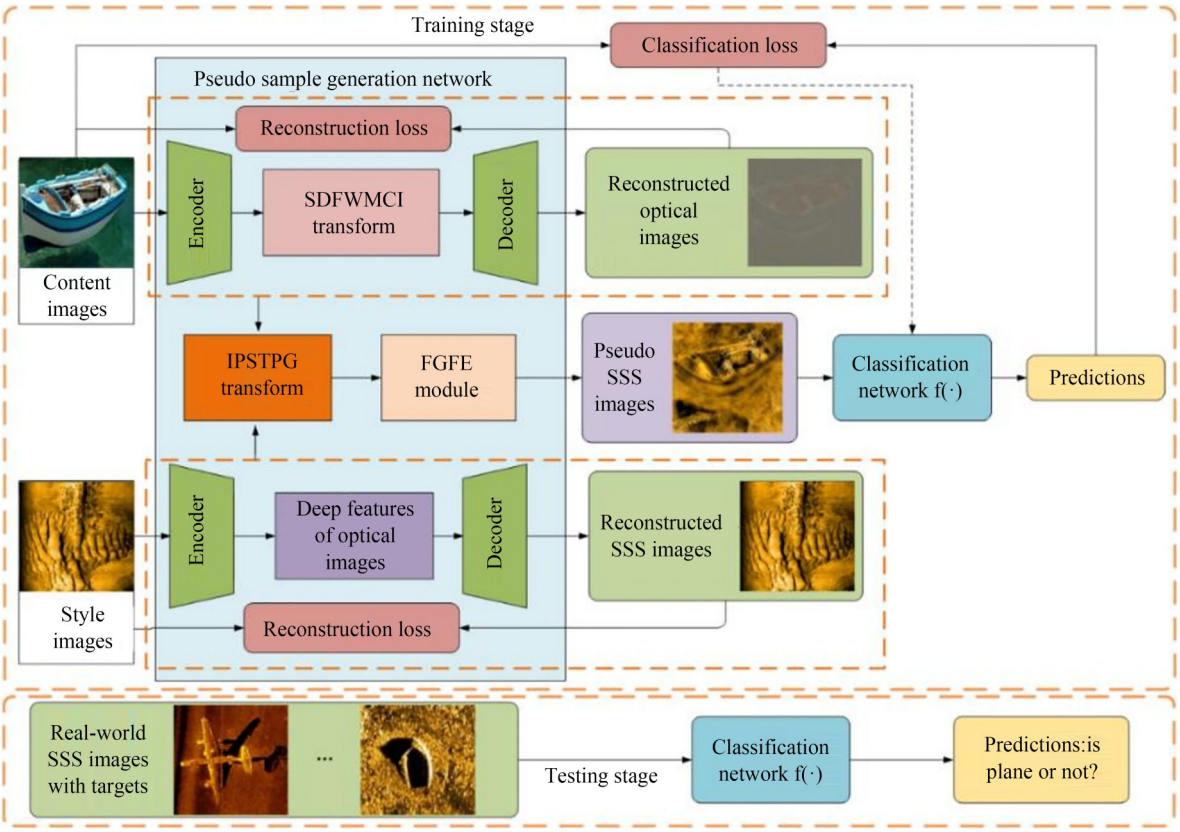


图2　本文方法的网络架构图

Fig. 2　Network architecture diagram of this method

2.1　内容图像特定领域特征削弱模块设计

SSS图像与传统光学图像相比在像素分布和图像纹理等特征方面存在较大差异。研究发现，目标在经过侧扫描声呐仪器成像后发生了改变，但与对应的光学图像相比，SSS成像物体的轮廓特征基本保持不变。本研究选择将这一不会变动的轮廓特征定义为不变领域特征，将颜色、像素等纹理特征定义为特定领域特征。在保持轮廓特征不变的情况下，将光学图像作为内容源域，真实SSS图像作为目标域，实现特征削弱。其具体过程分为6个步骤：(1)使用预训练的自编码器提取输入图像特征，使用反池化替换原上采样；(2)对光学内容图像进行编码，获得深度特征，采用白化变换去除内容图像的风格信息，保留全局内容结构；(3)使用Haar小波对白化后的深度特征进行分解；(4)在白化的深度特征上加入噪声，削弱输出解码器的能力；(5)在解码器的不同层级使用上采样和反池化；(6)利用像素重建损失和特征损失重构输入图像。通过公式描述：

$$y_p = f\left(s\left(e\left(x_j\right)\right)\right), x_j \in X_{oc}, X_{ss}, \tag{1}$$

其中：x_j是削弱模块的输入；$e(\cdot)$为特征提取网络；$s(\cdot)$是负责特征分离的函数；$f(*)$是实现深度特征与预测之间映射的分类器。X_{oc}为光学图像的实例集，X_{ss}表示SSS图像的实例集。在训练阶

段，首先预训练自编码器，在编码器-解码器训练过程中利用像素重建损失[26]和特征损失[27]对输入图像重新构建。

$$L_{\mathrm{R}}(\omega)=\|x_{\mathrm{R}}-x_i\|_2^2+\lambda\|e(x_{\mathrm{R}})-e(x_i)\|_2^2,\quad(2)$$

其中：x_{R}表示重构输出，x_i表示输入图像；$e(x)$是提取图像深度特征的编码器；λ表示权重参数。特征削弱任务在自编码器网络训练完成后开始，使用预训练的自编码器提取输入的内容图像特征，且不再使用$\|e(x_{\mathrm{R}})-e(x_i)\|_2^2$来约束输出特征。当$\boldsymbol{f}_{\mathrm{oc}}$已知时，需对$\boldsymbol{f}_{\mathrm{oc}}$减去其均值向量$\boldsymbol{m}_{\mathrm{oc}}$进行中心化操作，然后进行如公式(3)白化变换。

$$\hat{\boldsymbol{f}}_{\mathrm{oc}}=\boldsymbol{E}_{\mathrm{oc}}\boldsymbol{D}_{\mathrm{oc}}^{-\frac{1}{2}}\boldsymbol{E}_{\mathrm{oc}}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{oc}},\quad(3)$$

其中：$\boldsymbol{D}_{\mathrm{oc}}$是以特征值为$\boldsymbol{f}_{\mathrm{oc}}\boldsymbol{f}_{\mathrm{oc}}^{\mathrm{T}}$的协方差矩阵的对角矩阵，$\boldsymbol{E}_{\mathrm{oc}}$是相对应于特征向量的正交矩阵。$\hat{\boldsymbol{f}}_{\mathrm{oc}}$表示白化的内容图像特征，$\boldsymbol{f}_{\mathrm{oc}}$是内容图像深度特征。其次，通过着色变换将编码器从风格图像中提取的风格特征与白化的内容特征融合，用公式(4)表达：

$$\hat{\boldsymbol{f}}_{\mathrm{cs}}=\boldsymbol{E}_{\mathrm{ss}}\boldsymbol{D}_{\mathrm{ss}}^{\frac{1}{2}}\boldsymbol{E}_{\mathrm{ss}}^{\mathrm{T}}\hat{\boldsymbol{f}}_{\mathrm{oc}},\quad(4)$$

其中：$\hat{\boldsymbol{f}}_{\mathrm{cs}}$表示着色特征，$\boldsymbol{D}_{\mathrm{ss}}$是以特征值为$\boldsymbol{f}_{\mathrm{ss}}\boldsymbol{f}_{\mathrm{ss}}^{\mathrm{T}}$的协方差矩阵的对角矩阵，$\boldsymbol{E}_{\mathrm{ss}}$是相对应于特征向量的正交矩阵，$\boldsymbol{f}_{\mathrm{ss}}$是风格图像深度特征。最后将合并的深度特征$\hat{\boldsymbol{f}}_{\mathrm{cs}}$由解码器解码，重构出图像$x_{\mathrm{R}}$。

在本次任务中，当仅保留轮廓特征时，风格转换效率可以提高。若直接用该白化变换作为特征削弱网络，重构出的图像仍有残留的风格特征以及部分颜色，如图3(b)所示。为了去除多余的特征，本文对白化的深度特征进行小波变换分解，分解后可以得到低频分量和高频分量。在解码器的不同深度层级交替使用上采样和反池化操作。同时采用最近邻插值对二者长度进行修复，并通过解码器重构高低频特征图。真实的SSS图像细节是不够丰富的，但从图3(c)中可知，解码重建后的低频特征图像颜色和纹理非常清晰(彩图见期刊电子版)，细节过于丰富，会影响生成图像的质量，这并非本研究所需要的。本文提出了使用随机噪声代替部分非极大值的上采样方法，该方法是通过反池化层实现的，并非将每个非极大值都为零。如图3(d)所示，解码后高频特征已削弱绝大部分的风格特征以及颜色，保留了图像的主要结构，同时得到了较好的白化效果。

图3　深度特征解码重建结果

Fig. 3　Depth feature decoding and reconstruction results

虽使用反池化层能够较好保留原图结构信息，也会还原输入图像中的纹理细节。为解决这个问题，本文在适当的位置使用上采样层，上采样层以忽略活性位置的方式扩大了特征图，这可能会破坏特征信息的传输，这正是本文需要的。上采样层和反池化层的操作过程如图4所示。本文的编码器部分借鉴了PhotoWCT方法[16]中的设计，对于编码器来说，浅层通常会提取局部细节纹理和颜色特征，深层则提取更抽象的信息，如轮廓和大小。对于与解码器对称的编码器，浅层通常重构抽象的全局轮廓特征，而较深层则重构细节纹理和颜色特征。因此在解码器设计上，本文针对纹理信息抑制任务在解码器结构进行了改进，采用“浅层采用反池化层、深层使用上采样层”的策略，在保持主要轮廓还原能力的前提下，尽可能削弱纹理。本文的内容图像特定领域特征削弱模块结构如图5所示。

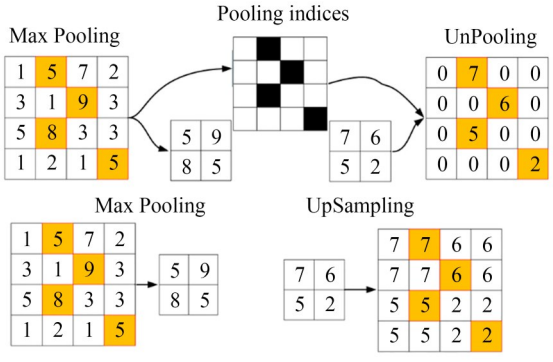


图4　上采样和反池化层的操作

Fig. 4　Operation of upsampling and unpooling layers

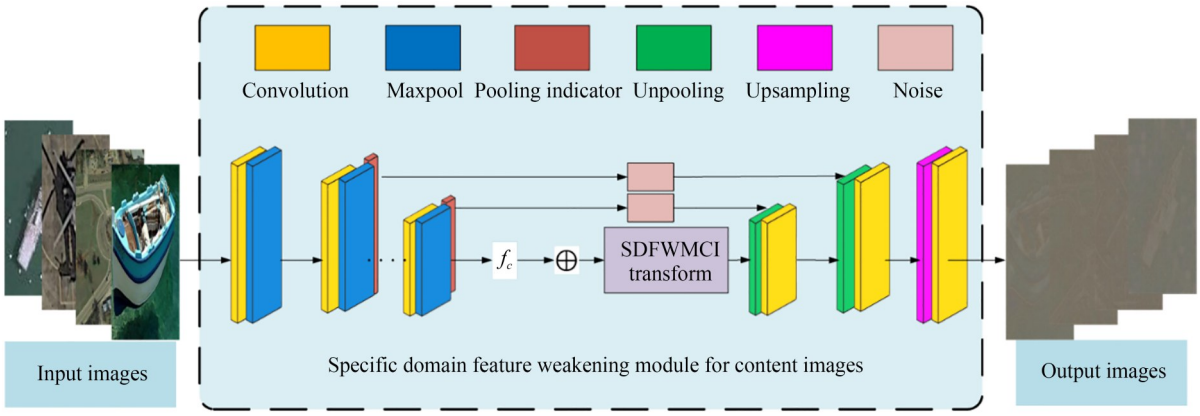


图 5　特征削弱方法的结构图

Fig. 5　Structure diagram of feature reduction method

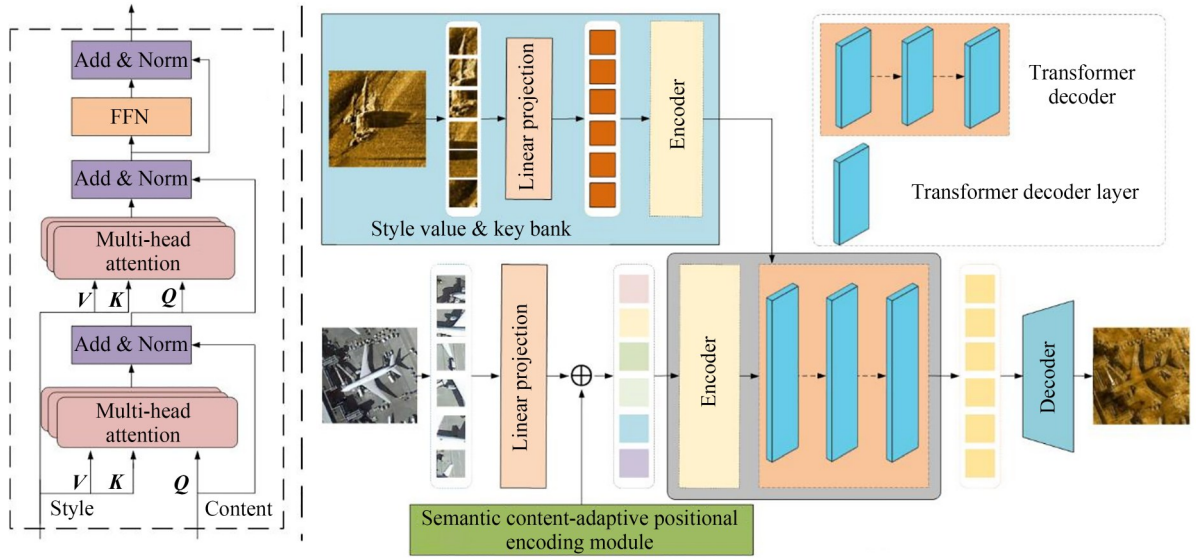
2.2　基于图像块的风格迁移伪图像生成方法设计

本文采用风格迁移方法生成伪图像。首先将光学内容图像和侧扫描声呐风格图像分割成若干图像块,并通过线性投影得到内容和风格图像块的序列。将设置了语义内容自适应位置编码模块(Semantic Content-adaptive Positional Encoding module,SCPE)的内容图像块序列输送到内容图像编码器,将风格图像块序列输送到风格图像编码器。两种图像块序列分别经过各自编码器编码处理后,运用多层解码器对内容图像块序列进行声呐图像风格化处理。最后使用解码器来获得伪 SSS 图像。基于图像块的伪图像生成模型结构图,如图 6(b)所示。

在二维条件下,像素(x_j, y_j)和像素(x_l, y_l)两个位置的图像块之间存在的相对位置关系为:

$$\boldsymbol{P}(x_j, y_j)^{\mathrm{T}}\boldsymbol{P}(x_l, y_l)=\sum_{k=0}^{(d/4)-1}\left[\cos\left(w_k(x_l-x_j)\right)+\cos\left(w_k(y_l-y_j)\right)\right], \tag{5}$$

其中:设置$w_k=(1/10\,000)^{2k/128}$,$d=512$。



(a) 表示transformer 解码器层的结构图
(a) Shows the structure of transformer decoder layer

(b) 表示生成模型示意图
(b) Shows the schematic diagram of the generated model

图 6　伪图像生成模型结构图

Fig. 6　Structure of the pseudo image generation model

本文采用$d(\cdot,\cdot)$表示两个图像块之间的空间距离。如图7所示为语义内容自适应位置编码模块示意图。在风格迁移中最好的结果就是相似的内容图像块风格化的效果也应该相似，即图7左边红色图像块分别于绿色、蓝色图像块之间风格化的效果相似(彩图见期刊电子版)。本文提出了语义内容自适应位置编码模块，来适应图像大小的调整。SCPE与传统的位置编码不同，SCPE是以图像的语义内容为条件。调整图像的大小不会影响两个图像块之间的空间关系。

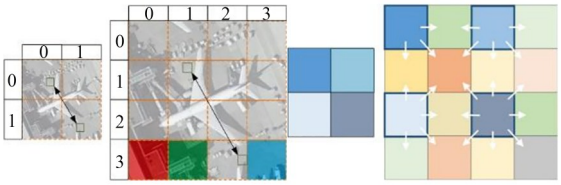


图7　语义内容自适应位置编码模块示意图

Fig. 7　Schematic diagram of semantic content-adaptive positional encoding module

将图像块(x,y)的SCPE记作$\boldsymbol{P}_{\mathrm{SC}}(x,y)$：

$$\boldsymbol{P}_L=F_{\mathrm{pos}}(AvgPool_{n\times n}(\varphi)),\tag{6}$$

$$\boldsymbol{P}_{\mathrm{SC}}(x,y)=\sum_{e=0}^{N}\sum_{f=0}^{N}(a_{\mathrm{ef}}\boldsymbol{P}_L(x_e,y_f)),\tag{7}$$

其中：$AvgPool_{n\times n}$表示平均池化，F_{pos}表示位置编码函数，$\boldsymbol{P}_L$是图像块序列φ的可学习位置编码，a_{ef}表示插值权重，N为相邻的图像块数。

生成网络有两个transformer编码器进行特征编码，用于图像块序列的领域转换。输入内容图像块序列$Z_{\mathrm{oc}}=\{\varphi_{\mathrm{oc1}}+\boldsymbol{P}_{\mathrm{SC1}},\varphi_{\mathrm{oc2}}+\boldsymbol{P}_{\mathrm{SC2}},\cdots,\varphi_{\mathrm{oc}L}+\boldsymbol{P}_{\mathrm{SC}L}\}$，并将该图像块序列送入内容编码器。将输入的图像块序列进行编码，将其转换为查询($\boldsymbol{Q}$)、键($\boldsymbol{K}$)和值($\boldsymbol{V}$)，编码操作后的内容图像块序列T_{oc}：

$$T'_{\mathrm{oc}}=F_{\mathrm{MSA}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})+\boldsymbol{Q},\tag{8}$$

$$T_{\mathrm{oc}}=F_{\mathrm{FFN}}(T'_{\mathrm{oc}})+T'_{\mathrm{oc}},\tag{9}$$

其中，$F_{\mathrm{FFN}}(T'_{oc})=\max(0,T'_{oc}W_1+b_1)W_2+b_2$。

输入风格图像块序列$Z_{ss}=\{\varphi_{\mathrm{ss1}},\varphi_{\mathrm{ss2}},\cdots,\varphi_{\mathrm{ss}L}\}$，以相同的操作使其被编码成风格图像块序列$T_{\mathrm{ss}}$，但与内容图像块序列$T_{\mathrm{oc}}$不同的是该序列没有应用位置编码操作。解码器是以风格图像块序列T_{ss}为根基，对内容图像块序列T_{oc}进行回归式的解码操作，每个元素能够逐步捕捉并适应风格序列T_{ss}的特征。图6(a)为transformer解码器层的结构图。本次使用的回归式解码是对图像块序列采用一次性全部输入的方式进行预测输出。解码器的输出G的计算过程：

$$G''=F_{\mathrm{MSA}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})+\boldsymbol{Q},\tag{10}$$

$$G'=F_{\mathrm{MSA}}(G''+\boldsymbol{P}_{\mathrm{SC}},\boldsymbol{K},\boldsymbol{V})+G'',\tag{11}$$

$$G=F_{\mathrm{FFN}}(G')+G',\tag{12}$$

其中：运用图像块序列T_{oc}生成查询$\boldsymbol{Q}$，通过风格图像块序列T_{ss}生成键$\boldsymbol{K}$、值$\boldsymbol{V}$。Transformer的输出为$(HW/64)\times C$的形状，本研究没有直接进行上采样来构建最终结果，最后，采用三层的CNN解码器来优化图像块序列输出结果。在每一层中，使用了能够扩大规模的操作(包括3×3Conv+ReLU+2×Upsample)。最终，可以得到$H\times W\times 3$的最终结果。如图8所示，采用Transformer解码器输出结果存在细节模糊伪影等不足，经CNN解码器改进后输出的初始伪图像具有更加清晰地目标边缘轮廓。

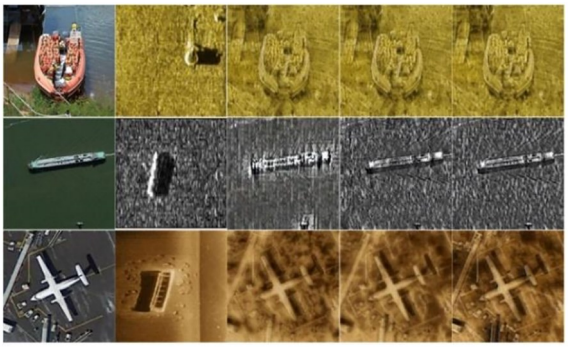


图8　IPSTPG网络优化后输出及引导增强结果图

Fig. 8　Output and guidance enhancement results after IP-STPG network optimization

为了使生成图像能够更好地兼具原目标的结构和声呐风格特征。本文引入了两种感知损失，即内容损失和风格损失[19,21]。内容损失$\boldsymbol{L}_c$的计算过程：

$$\boldsymbol{L}_c=\frac{1}{N_l}\sum_{j=0}^{N_l}\left\|\phi_j(I_P)-\phi_j(I_{\mathrm{OC}})\right\|_2,\tag{13}$$

其中：$\phi_j(\cdot)$为在第j层提取的特征，N_l表示层数，I_P表示初始伪图像。

风格损失$\boldsymbol{L}_s$的计算过程：

$$L_s=\frac{1}{N_l}\sum_{j=0}^{N_l}\left\|\alpha(\phi_j(I_P)-\phi_j(I_{SS}))\right\|_2+\left\|\beta\left(\phi_j(I_P)-\phi_j(I_{SS})\right)\right\|_2,\tag{14}$$

其中：$\alpha(\cdot)$表示提取特征的均值，$\beta(\cdot)$对应提取特征的方差。

此外，本文还引入了身份损失[28]来生成更加自然、和谐的内容和风格表达。分别将一对相同的风格图像I_{SS}或内容图像I_{OC}作为IPSTPG网络的输入，可以得到图像I_{PS}或图像I_{PC}。身份损失计算示意图，如图9所示。由于使用相同的图像作为输入，身份损失使模型能够专注于保留图像本身的内容结构，而不仅仅是风格转换。因此，无论是来自光学图像的结构，还是来自声学图像的风格特征，都可以得到更有效地保留。

$$L_{id1}=\|I_{PC}-I_{OC}\|_2+\|I_{PS}-I_{SS}\|_2,\tag{15}$$

$$L_{id2}=\frac{1}{N_l}\sum_{j=0}^{N_l}\left\|\phi_j(I_{PC})-\phi_j(I_{OC})\right\|_2+\left\|\phi_j(I_{PS})-\phi_j(I_{SS})\right\|_2.\tag{16}$$

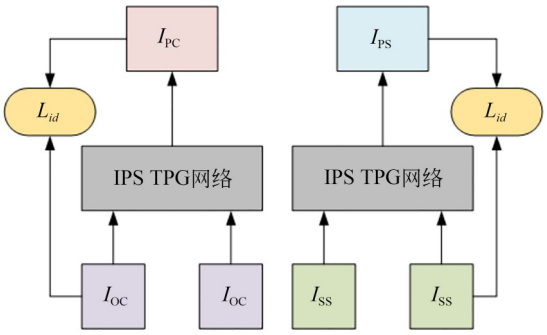


图9　身份损失计算示意图

Fig. 9　Schematic diagram of identity loss calculation

IPSTPG的总损失L：

$$L=\lambda_c L_c+\lambda_s L_s+\lambda_{id1}L_{id1}+\lambda_{id2}L_{id2},\tag{17}$$

其中：λ_c，λ_s，λ_{id1}，λ_{id2}是用于平衡各类损失的权重参数。

2.3　快速引导滤波增强模块设计

由于本文加入了随机噪声来削弱内容图像特征，同时借助该噪声模拟水下噪声环境。噪声对生成图像的影响是不可回避的，且会降低生成图像的视觉效果。生成的伪图像获得干净的可参考真实图像是颇具难度或者不可信的，因此无法使用干净图像作为目标对噪声图像进行增强操作。本文引入了一种快速引导滤波增强模块，能够平衡生成图像中噪声。

快速引导滤波包含引导图像I_g（表示需要增强的初始伪SSS图像本身），滤波输入图像I_P，滤波输出图像的局部线性模型q[29-30]。引导滤波器及其核的定义如式(18)：

$$q_j=a_l I_{gj}+b_l,\forall j\in\omega_l,\tag{18}$$

其中：j表示像素在引导滤波器中的索引，ω表示半径为r的局部窗口，l表示窗口ω的索引。a_l，b_l表示窗口ω_l中的线性系数。在给定图像I_P输入的条件下，由上式(18)可以得到：

$$a_l=\frac{\frac{1}{|\omega|}\sum_{j\in\omega_l}I_{gj}I_{pj}-\mu_l\overline{I_{pl}}}{\sigma_l^2+\varepsilon},\tag{19}$$

$$b_l=\overline{I_{pl}}-a_l\mu_l,\tag{20}$$

其中：σ_l，μ_l为引导图像I_g在窗口l中的均值、方差，ε表示与平滑程度相关的正则化参数。输出q表达式：

$$q_j=\frac{1}{|\omega|}\sum_{l:j\in\omega_l}(a_l I_{pj}+b_l)=\overline{a_j}I_{gj}+\overline{b_j},\tag{21}$$

其中：$\overline{a_j}$，$\overline{b_j}$表示所有窗口的平均系数，且窗口是以j为中心。当方差较小时，且引导图为输入图像本身，则输出图像为进行滤波处理后的图像，实现滤波效果。当方差较大时，实现保边效果。

然而，传统的引导滤波器依赖引导图像，在去噪时无法实现快速计算。在快速引导滤波中通过在图像上计算线性系数，再进行上采样恢复原始分辨率，可将计算复杂度从$O(E)$降至$O(E/s^2)$，s表示下采样比例，E表示像素数量。过大的下采样比例s会造成系数估计的偏差累积，从而影响最终输出图像的边缘对齐精度与局部一致性。当s较大时，为了抵消下采样带来的精度损失，往往需要减小正则化参数ε或增大局部窗口半径r以稳定局部细节；半径r决定了图像中每个像素在进行局部线性拟合时所参考的邻域范围，较小的r值有助于增强局部结构，尤其是边缘的清晰度与准确性。相反，较大的r值能显著提升图像的整体平滑性，但也可能引入边缘模糊的问题；正则化参数ε能够控制局部线性模型的稳定性，较小的ε有助于增强边缘细节，但容易放大微小噪声或产生伪影；而较大的ε能够增强平滑效果，但边缘与细节信息会相应丢失。因此，在本研究应用中，遵循文献[29]的经验对参数

进行设置，$s=4$，$r=16$，$\varepsilon=0.01$，以实现具备良好边缘保持特性的滤波效果。如图 8 所示，FGAE 表示增强结果，初始伪图像经过 FGFE 增强后噪声阴影有所减少，目标轮廓边缘更加清晰。

3　实验结果与分析

为了确保实验的公正性、严谨性，本文参考文献[24]预训练思路，选用了 MS-COCO[31]作为内容数据集进行预训练，同时使用 WikiArt[32]作为风格数据集进行预训练。实验在 NVIDIA GeForce RTX 4090 GPU 上进行。在生成网络设置中，所有图像都以 512×512 进行加载，并随机裁剪为 256×256 进行数据增强，批大小设置为 8，使用预热调整策略将学习率设置为 0.000 5，使用 Adam 优化器优化参数，使用 16 万次迭代来训练。

3.1　数据集描述

对于伪 SSS 图像生成，本次研究收集了 639 幅 SSS 图像，总计 14 类。从中选择合适的图像作为生成伪 SSS 图像的风格图像数据集，定义 SSS 风格图像数据集为 S_{SI}，并与任务[13,24]中使用的风格数据集对齐，其中包括五类 SSS 目标图像，共 43 张用于本次任务。表 1 提供了数据集各类型的信息。

由于 SSS 仪器成像角度和距离的特性，该声呐图像通常是俯视图，风格迁移是通过参考内容生成的图像，在生成伪 SSS 图像时，需考虑其与真实 SSS 图像的视角相似性。因此，选择航拍遥感图像作为内容图像，以实现风格上的匹配，选用与文献[14,24]相同俯视视角的光学遥感数据集，从中挑选部分图像，构建并定义新光学内容图像数据集为 O_{CI}。内容图像包括飞机（1 313 张）、轮船（1 019 张）和其他（2 000 张）三类。采用预训练的 ResNet-152[33]模型作为 SSS 图像分类器。为评价风格迁移的效果，采用与工作[24]相同的真实彩色 SSS 样本数据集作为测试数据集 U_t，数据集 U_t 中包含 34 架飞机、240 艘沉船。为了消除伪彩色的干扰，本研究将灰度的真实 SSS 图像数据集 KLSG[34]作为测试数据集 U_v，来验证本文方法的优越性，KLSG 数据集具有飞机声呐图像 62 张、沉船声呐图像 385 张。

表 1　数据集各类型的信息

Tab. 1　Information about each type of dataset

Datasets	Classes	Number	Total
O_{CI}	Plane	1 313	4 332
	Ship	1 019	
	Others	2 000	
S_{SI}	Body	6	43
	Mine	10	
	Pipeline	4	
	Ship	12	
	Others	11	
U_t	Plane	34	274
	Wreck	240	
U_v	Plane	62	447
	Wreck	385	

3.2　评估指标

本研究引入评价图像生成模型生成声呐图像质量的指标主要有风格损失、学习感知图像块相似度（Learned Perceptual Image Patch Similarity，LPIPS）[35]，内容损失、结构相似度（Structural Similarity，SSIM）[36]对生成图像的风格化质量进行评估。其中，学习感知图像块相似度是一种用于衡量图像相似性的指标，更接近人类视觉系统的感知，通过学习图像块之间的相似性，从而能够捕捉到人类视觉系统所关注的细节。LPIPS 公式如式（22）：

$$d(x,x_0)=\sum_l \frac{1}{H_l W_l}\sum_{h,w}\|w_l\odot(\hat{y}^l_{hw}-\hat{y}^l_{0hw})\|_2^2, \quad (22)$$

其中：d 为 x 与 x_0 之间的距离，x 与 x_0 分别表示生成图像和参考图像，l 表示预训练的深层网络的第 l 层；$\hat{y}^l$，$\hat{y}^l_0$ 分别表示图像 x，x_0 在深层网络第 l 层输出的特征图在通道维度上进行单位范数归一化后的结果；H_l，W_l 表示第 l 层特征图的空间尺寸（高度与宽度），w_l 表示第 l 层的通道加权向量。

结构相似度主要从图像的亮度、对比度以及结构信息衡量图像的结构相似性，从而量化生成图像保留内容图像的结构信息的能力，其值越大说明结构维持越好。其原理如式（23）：

$$\mathrm{SSIM}=\frac{(2\upsilon_x\upsilon_y+Z_1)(2\xi_{xy}+Z_2)}{(\upsilon_x^2+\upsilon_y^2+Z_1)(\xi_x^2+\xi_y^2+Z_2)}, \quad (23)$$

其中：υ_x 和 υ_y 分别表示输入内容图像和生成图像

的均值，υ_x^2和υ_y^2表示对应的方差，ξ_{xy}为协方差，Z_1，Z_2表示常数。为了验证本研究的性能提升并非偶然，引入了t-test显著性检验[37]。

$$t=\frac{\bar{H}_1-\bar{H}_2}{\sqrt{\dfrac{(a_1-1)S_1^2+(a_2-1)S_2^2}{a_1+a_2-2}\left(\dfrac{1}{a_1}+\dfrac{1}{a_2}\right)}},\tag{24}$$

其中：t表示统计量，$\bar{H}_1$和$\bar{H}_2$分别表示本文方法数据和对比方法数据的均值，S_1^2，S_2^2分别表示本文和对比方法数据的方差，a_1，a_2分别表示对应的样本容量。设置α为0.05。

3.3　与SOTA方法比较

在本章节中将进行多种实验方法的对比，用以证明本文方法的在零样本SSS图像的分类任务中的优越性，包括生成伪SSS图像的视觉质量、目标风格化的效果、分类的性能以及模型效率对比。

3.3.1　视觉对比

为了验证所提伪图像生成网络的先进性，本文使用WCT[15]，AdaIN[21]，PhotoWCT[16]，AAMS[17]，MCC[18]，ArtFlow[19]，IEST[20]这7种先进的风格迁移方法进行生成图像视觉比较。

为了提高实验的严谨性，本文采用了多种类别的内容目标图像，包括飞机、船舶、地面道路。同样采用了不同目标类别的声呐风格图像，包括海底地势、沉船、矿石等。如图10所示（彩图见期刊电子版），WCT方法采用全局特征统计变换实现风格迁移，然而其基于协方差矩阵的内容特征处理机制会导致局部结构特征的退化。（例如，WCT列中的(a)和(c)）。由于AdaIN算法生成的图像出现目标的细节特征缺失，风格化效果欠佳（例如，AdaIN列中的(b)和(d)）。AAMS关注的是内容图像的主体结构，而忽略了其他部分。导致模型生成的图像风格化不充分，残留了部分颜色干扰（例如，AAMS列中的(b)和(e)）。PhotoWCT风格迁移的效果较好，但存在不合理噪声分布，且清晰度较低（例如，PhotoWCT列中的(e)和(f)）。MCC方法生成的伪SSS图像存在失真的现象（例如，MCC列中的(d)和(f)）。ArtFlow模型通过优化图像重建误差和恢复偏差来实现风格迁移，但基于流的架构在特征提取方面存在局限性，导致生成图像的边缘区域出现渲染失真现象（例如，ArtFlow列中的(c)和(f)）。IEST方法通过整合非局部注意力机制与对比学

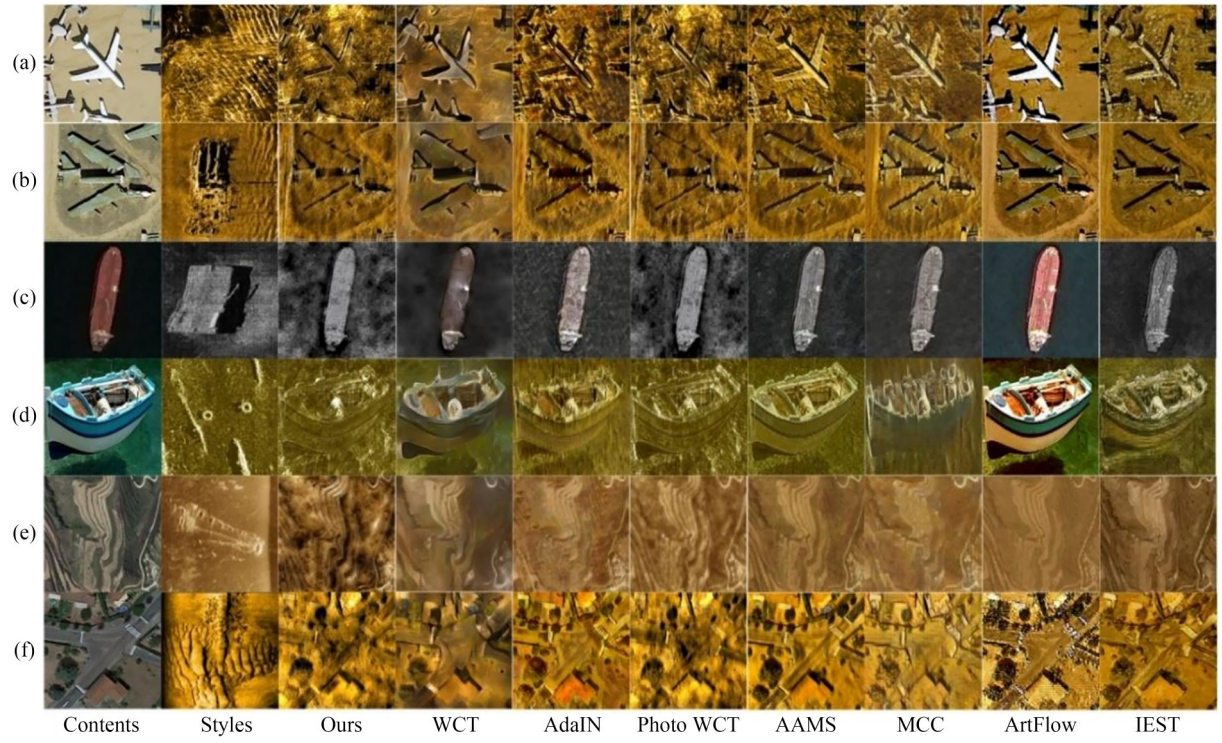


图10　本文方法与7种风格迁移方法的结果及对比

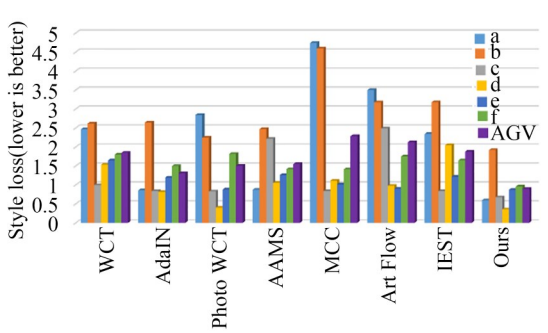
Fig. 10　Results and comparison between this method and seven style transfer methods

习策略,有效挖掘风格化图像间的潜在特征。但该方法在全局风格特征建模方面仍存在不足,导致整体风格一致性有待提高。(例如,IEST 列中的(a)和(d))。结合以上几种方法的对比结果可知,本文的方法在模拟 SSS 图像目标的外部细节特征,在处理声学图像和光学图像两个不同领域之间的潜在相关性方面具有显著效果(例如,Ours 列中的(d)和(e))。

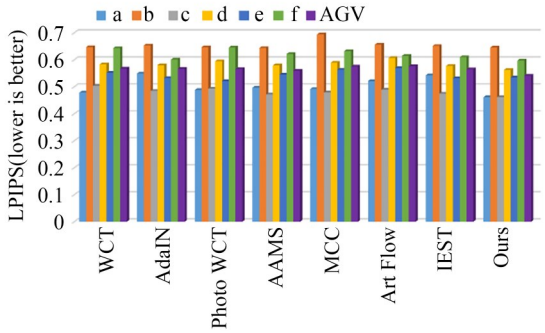
3.3.2　风格化渲染质量对比

通过计算风格迁移渲染后的输出图像 I_{CS} 与用于风格迁移输入的内容图像和风格图像之间的差异,进行评估风格迁移方法的先进性。本文基于公式(14)计算各组图像的风格化差异,并引入学习到的感知图像块相似度作为另外一种评价指标。为了实验的严谨性,本文基于公式(13)和结构相似度计算图像的内容信息方差,对各组方法在内容图像光学结构的保留程度进行评估。如图 11 所示,风格化渲染质量评价指标结果((a)~(f)对应图 10 中的 6 组图像、AGV 对应各评价指标的平均值)。

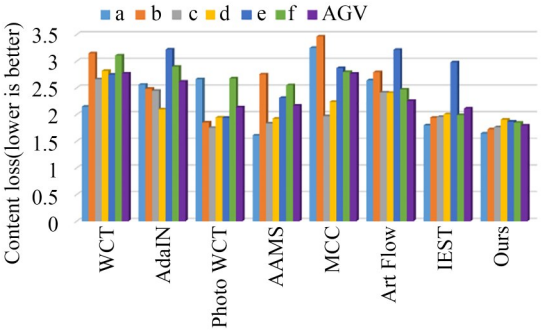
在图 11 中展示了本文方法与七种先进方法的风格化指标比较,其中采用了 L_s,LPIPS,L_c,SSIM 四个评价指标,对图 10(a)~(f)中的 6 组伪 SSS 图像进行风格化效果评估。在图 11(a)所示的 L_s 中,IPSTPG 方法在(a)~(f)组得分为 0.587 1,1.923 9,0.674 4,0.359 7,0.872 6,0.962 1,在 L_s 上的平均得分为 0.896 6,相较次优方法提升了 0.413 0。在指标 LPIPS 方面,如图 11(b),IPSTPG 方法在(a)~(f)组得分为 0.458 9,0.648 3,0.464 1,0.572 0,0.528 5,0.593 1,仅在(b)组和(e)组上分别落后 AAMS 算法、PhotoWCT 算法。IPSTPG 方法的 LPIPS 平均分为 0.544 1,相较次优方法提升了 0.019 9。在指标 L_c 方面,如图 11(c),IPSTPG 方法在(a)~(f)组得分为 1.640 4,1.718 2,1.759 6,1.905 1,1.871 0,1.838 4,仅在(a)组、(c)组上分别落后 AAMS 算法、PhotoWCT 算法。在 L_c 上的平均分为 1.788 8,相较次优方法提升了



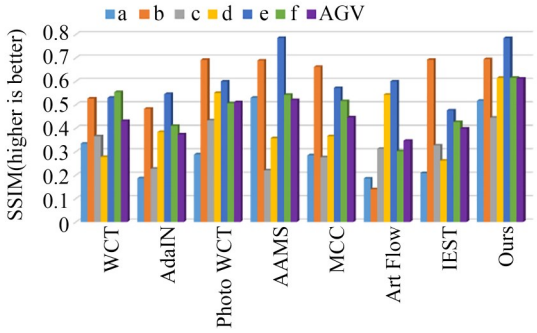
(a) 风格损失值的对比柱形图
(a) Comparison bar graphs of style loss values



(b) LPIPS值的对比柱形图
(b) Comparison bar graphs of LPIPS values



(c) 内容损失值的对比柱形图
(c) Comparison bar graphs of content loss values



(d) SSIM值的对比柱形图
(d) Comparison bar graphs of SSIM values

图 11　各模型生成图像的四种评价指标的对比柱形图

Fig. 11　Comparative bar graphs of four evaluation indicators of images generated by each model

0.320 6。在指标SSIM方面，如图11(d)，IPST-PG方法在(a)～(f)组得分为0.518 0，0.694 7，0.459 6，0.621 1，0.783 2，06 217，在(a)组和(e)组得分略低于AAMS。但本方法的SSIM平均分为0.616 2，相较次优方法提升了0.091 3。根据以上对比实验可以得出，本文所提方法生成的伪图像更加接近真实SSS图像，且风格化质量更高，有效保留了光学目标的大量细节特征。综上所述，本方法能够有效地平衡光学内容的细节特征和声呐风格特征的分布，在保留光学目标结构的同时渲染出高质量的具有SSS风格的图像。

3.3.3　分类性能对比

将上述风格迁移方法生成的伪图像用于ResNet-152分类器训练，然后对未使用的真实SSS图像数据集进行分类。对本文方法和7种对比方法的全局精度和平均精度进行合理评估。以所有真实的SSS样本图像的识别精度表示全局精度；以真实的SSS样本图像的平均每类识别精度表示平均精度。如表2所示，本文方法在数据集U_t的飞机残骸34个样本中精准识别分类23个，在沉船240个样本中精准识别分类210个。从表2中，本文所用方法在彩色数据集U_t上得出的全局准确率高达85.03%，与次优方法相比增加了4.74%。平均准确率高达77.58%，与次优方法相比增加了10.28%。根据以上实验可以得出，本方法能够生成SSS图像分类所需的高质量伪SSS样本，实现零样本SSS图像分类性能的显著提升。

由于真实的SSS图像是不存在颜色渲染的，为了更好摆脱颜色在SSS图像分类任务中可能存在的干扰，本文采用灰度光学图像和灰度声呐图像生成伪图像再次训练ResNet-152。如表3所示，采用灰度数据集U_v分别对7种SOTA方法进行分类验证。分类实验结果显示，本文方法在灰度数据集U_v的飞机残骸62个样本中精准识别分类42个，在沉船385个样本中精准识别分类344个。本文所提方法实现了86.35%的全局准确率，在平均准确率方面，准确率达到了78.54%。根据以上实验结果证明，即使在处理RGB单通道图像的情况下，本文的方法相较于其他方法也显示出更先进的性能。这进一步验证了本文方法在实现声学特性转移方面的有效性。

为了验证本方法生成伪图像在各种分类模型上的普遍适用性，使用本文的伪SSS图像来训练多种分类模型，并在数据集U_t和U_v上进行验证，分类结果如表4所示。参考文献[24,38]的实验设置，选取验证的分类模型有VGG-19[39]，ResNet-152[33]，vision transformer-B/32(ViT-B/32)[40]，ViT-L/32[40]，EffcientNet-B6[41]，EffcientNet-B7[41]，MobileNetV2[42]和MobileNetV3[42]在彩色真实SSS图像数据集U_t的测试中，ResNet-152模型以85.03%的全局准确率表现最优。

表2　在数据集U_t上的全局精度和平均精度

Tab. 2　Global accuracy and average accuracy on dataset U_t

Methods	Plane total 34 samples	Wreck total 240 samples	Global accuracy	Average accuracy
WCT	22	125	53.65%	58.39%
AdaIN	**25**	133	57.67%	64.47%
PhotoWCT	17	203	80.29%	67.29%
AAMS	15	139	56.20%	51.01%
MCC	21	148	61.68%	61.71%
ArtFlow	4	158	59.12%	25.86%
IEST	19	104	44.89%	49.61%
Ours	23	**210**	**85.03%**	**77.57%**

其中，ViT-L/32模型在飞机类别的分类识别的效果最佳，但在沉船类别的分类识别的性能最低。如表4所示，ViT-L/32在U_t数据集上取得了最低的全局准确率80.66%和最高的平均准确率77.60%，而VGG-19网络的平均准确率表现最差。在灰度真实SSS图像数据集U_v上，EffcientNet-B6和EffcientNet-B7分别在平均精度和全局精度指标上取得最优结果，分别为81.64%

表 3　在数据集 U_v 上的全局精度和平均精度

Tab. 3　Global accuracy and average accuracy on dataset U_v

Methods	Plane total 62 samples	Wreck total 385 samples	Global accuracy	Average accuracy
WCT	37	278	70.47%	65.95%
AdaIN	**42**	211	56.60%	61.27%
PhotoWCT	38	296	74.72%	69.09%
AAMS	23	238	58.39%	49.45%
MCC	30	180	46.98%	47.57%
ArtFlow	5	337	76.51%	47.79%
IEST	26	245	60.63%	52.79%
Ours	**42**	**344**	**86.35%**	**78.54%**

表 4　多种分类模型在 U_t 和 U_v 上的全局和平均准确率

Tab. 4　Global and average accuracy of various classification models on U_t and U_v

Methods	Global accuracy on U_t	Average accuracy on U_t	Global accuracy on U_v	Average accuracy on U_t
VGG-19	83.21%	72.75%	80.76%	74.63%
ResNet-152	85.03%	77.57%	86.35%	78.54%
ViT-B/32	81.02%	76.55%	79.42%	75.20%
ViT-L/32	80.66%	77.60%	78.97%	74.94%
EffcientNet-B6	82.12%	73.38%	87.02%	81.64%
EffcientNet-B7	82.17%	75.91%	87.25%	81.10%
MobileNetV2	81.39%	74.23%	85.46%	76.00%
MobileNetV3	83.58%	75.48%	85.68%	76.81%

和 87.25%。根据分类结果可知，MobileNetV3 在数据集 U_t 上取得了次优的全局准确率 83.58%。在数据集 U_v 上，MobileNetV2 和 MobileNetV3 在平均准确率均优于 ViT-B/32 和 ViT-L/32。MobileNetV2 和 MobileNetV3 虽在网络体积、结构复杂度上无法与其他模型相比，但其在保证体积更小的同时依然取得较高的准确率，这为神经网络模型在嵌入式设备中运行提供了理论支持。实验结果表明，所有分类模型在 U_t 和 U_v 数据集上均展现出优异的分类性能，这不仅验证了本文方法生成的伪 SSS 图像与多种分类模型具有良好的兼容性，有效地证明了本文所提方法具有广泛的适用性。

本研究引入 t 检验对本文数据进行验证，证明本文方法显著优于其他对比方法。本研究在数据集 U_v 上进行检验，并分别对本文方法和 7 组对比方法进行 10 次重复分类实验记录数据全局和平均准确率。假设本文方法显著优于对比方法。表 5 中 A～G 表示将本文方法与其他方法逐个进行独立样本 t 检验。根据检验结果可知，在全部检验实验中 p 值均小于 0.05，且较大的 t 值表明两组均值之间的差异大于合并的标准误差，表明与对比组之间的差异显著，说明假设成立，本文方法领先对比方法，证明所提方法在提升准确率方面具有可靠性。

表 5　在数据集 U_v 上 t 检验的 t 值和 p 值

Tab. 5　T-values and p-values of t-test on U_v

Methods	Global accuracy		Average accuracy	
	t	p	t	p
A	46.42	4.998E-12	69.08	1.411E-13
B	392.39	2.308E-20	91.01	1.183E-14
C	171.16	4.035E-17	102.26	4.148E-15
D	226.28	3.272E-19	84.36	2.342E-14
E	368.38	4.074E-20	114.67	1.481E-15
F	53.71	1.352E-12	160.52	7.186E-17
G	143.59	1.958E-16	59.88	5.094E-13

3.3.4　效率比较

本研究的性能评估采用随机选取的 100 对内

容-风格图像（256×256和512×512分辨率）进行渲染速度测试。如表6所示，本方法在两种分辨率下分别达到0.032 s和0.063 s的平均处理时间，本文所提方法已被证明在不同分辨率下运行速度优于SOTA方法。对于模型参数，基于单层特征提取的方法，如AdaIN和PhotoWCT采用单层编码特征图来提取不同的图像特征，虽实现了更低的参数值，但牺牲了特征表达能力。因此，这些方法的模型参数超过了基于多层特征的方法。相比之下，本方法通过优化多层特征融合机制，在保持14.73 M适中参数量的同时，成功实现了高质量的SSS风格迁移。结果表明，本方法在计算效率和生成质量方面均优于对比方法，完全满足风格迁移的应用需求。

表6　两种分辨率下的平均处理时间以及各方法的参数

Tab. 6　Average processing time and parameters of each method at two resolutions

Methods	Time /s (256×256) ↓	Time /s (512×512) ↓	Parameter/M ↓
WCT	1.947	2.513	34.24
AdaIN	0.038	0.079	**7.11**
PhotoWCT	0.133	0.249	8.35
AAMS	2.082	2.204	49.46
MCC	0.069	0.153	17.26
ArtFlow	0.172	0.762	26.46
IEST	0.037	0.068	21.12
Ours	**0.032**	**0.063**	14.73

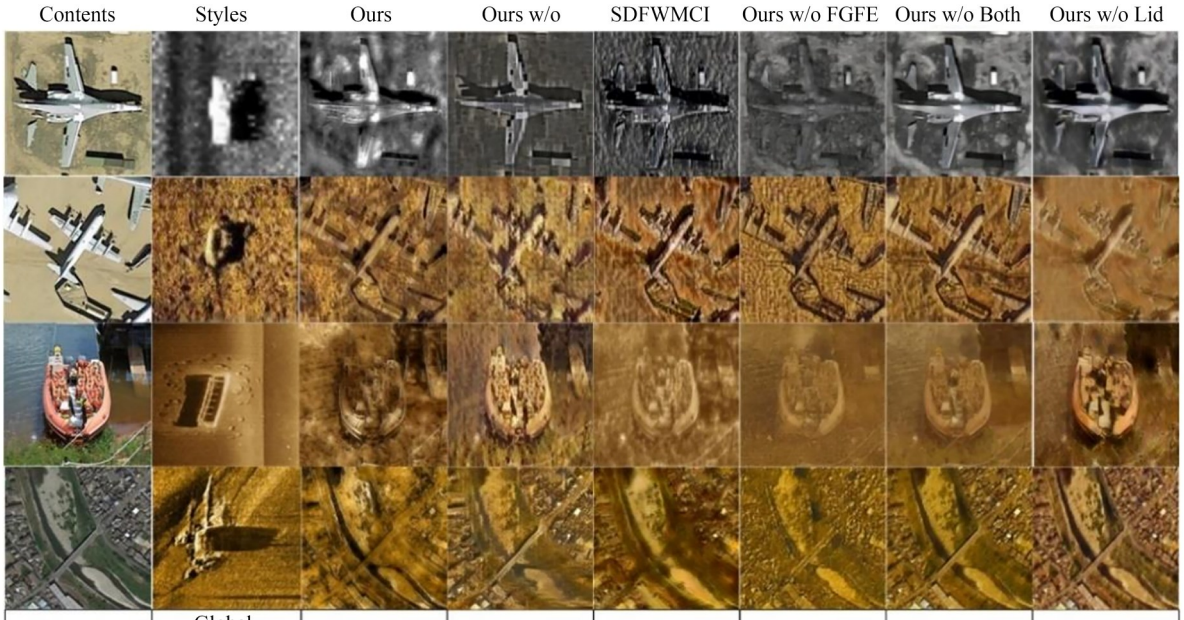
注：最优的结果用粗体表示

3.4　消融实验

在本章节中进行了几个消融实验来验证所提出的SDFWMCI设计和FGFE模块的性能。

3.4.1　SDFWMCI分析

本文提出了SDFWMCI设计，通过削弱光学内容图像中与SSS图像特征不相关的特征，来增强SSS风格特征与光学内容图像特征的融合。为了验证SDFWMCI在提高风格化质量方面的有效性，本文将特征削弱模块从网络中移除，并将其替换成InfoGAN[43]，β-VAE[44]经典解耦方法并将其命名为w/o SDFWMCI。图12的第三列表示添加SDFWMCI，第四列和第五列表示将



U_t	Global accuracy	85.03%	53.28%	60.58%	71.04%	53.57%	69.34%
	Averge accuracy	77.57%	53.14%	61.09%	59.23%	56.29%	64.83%
U_v	Global accuracy	86.35%	50.34%	59.73%	75.94%	58.51%	78.97%
	Averge accuracy	78.54%	45.46%	51.59%	68.65%	49.60%	69.53%

图12　SDFWMCI和FGFE模块的消融实验

Fig. 12　Ablation experiments of SDFWMCI and FGFE modules

SDFWMCI替换为InfoGAN，β-VAE模块后的渲染结果。根据实验结果可知，InfoGAN生成结果的风格化伪图像存在严重的棋盘效益和模糊失真（如第四列的第1行和第2行）。β-VAE当调整$\beta>1$时在质量上优于VAE（$\beta=1$），根据文献[44]设置$\beta=4$，其特征削弱结果的风格化图像如图12中第五列。尽管β值越大越有利于学习解耦表示，但代价是重建图像会出现模糊、变暗造成质量下降，导致生成的伪图像存在失真，且风格化结果的边缘轮廓辨识度低，与真实声呐图像存在差异（如第五列的第1，3，4行）。值得注意的是，添加了SDFWMCI的模型成功地去除了内容图像中的颜色信息。在U_t数据集上，全局精度和平均精度相比替换InfoGAN分别提高了31.75%，24.43%，与替换β-VAE相比分别提高了24.45%，16.48%。在U_v数据集上，全局精度和平均精度相比替换InfoGAN分别提高了36.01%，33.08%，与替换β-VAE相比分别提高了26.62%，26.95%。实验结果表明，SDFWMCI的应用有效地去除了特定领域特征对风格迁移的干扰，促进内容图像特征和SSS风格的融合，提高渲染结果的质量。

3.4.2 *FGFE*模块分析

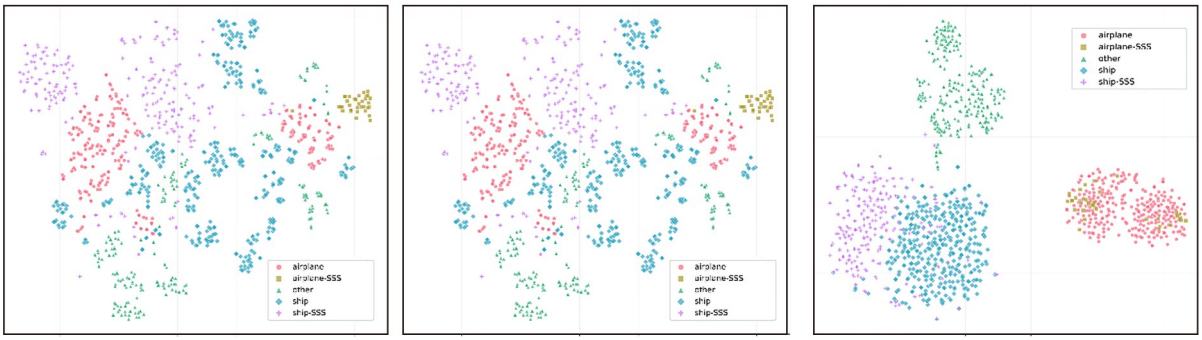
本文提出了FGFE模块，用来平衡风格迁移的伪SSS图像中的噪声分布，提高伪SSS图像的质量。为了验证FGFE模块在改善风格化质量方面的先进性，本文将快速引导滤波模块从网络中移除（w/o FGFE），将同时移除SDFWMCI、FGFE模块的网络命名为w/o Both，并进行风格化结果比较。图12的第三列、第六列和第七列，列出了添加FGFE模块、不添加FGFE模块、同时不添加SDFWMCI和FGFE模块时所提方法（w/o Both）的渲染结果。可以看出，第六列的w/o FGFE不能有效地平衡图像中的噪声分布，导致生成图像模糊，无法合理辨别目标（例如，w/o FGFE在第1组和第4组中的结果）。值得注意的是，添加了FGFE的模型成功地平衡了生成图像中噪声的分布。在U_t数据集上，全局精度和平均精度相比添加该模块前分别提高了13.99%、18.34%；在U_v数据集上，全局和平均精度相比增加前分别提高了10.41%和9.89%。实验结果表明，FGFE模块的应用有效地平衡了噪声的分布，合理模拟水下噪声干扰，生成更加逼真的SSS图像。

3.4.3 身份损失函数的分析

为了验证身份损失对伪图像生成和分类性能提升的有效性，本文提出对身份损失函数进行评估。本文将身份损失从网络在移除（w/o L_{id}），在图12中的第三列和第八列分别展示了添加身份损失和不添加身份损失的风格化渲染图片以及分类准确率。从实验结果表明，第八列w/o L_{id}的生成结果表现出视觉伪影、模糊和色彩干扰（如第2行、第3行），而带有身份损失的方法实现了光学内容结构的保留和声呐图像风格有效转移。图7中显示了两个模型在不同数据集上的识别精度。与w/o L_{id}相比，本文方法的全局和平均精度在数据集U_t上提高了15.69%和12.74%，在数据集U_v上提高了7.38%和9.01%。实验结果表明，身份损失函数在风格迁移任务中在内容结构的保持和风格特征的转移方面具有有效性，从而提升SSS图像生成的稳定性。

3.4.4 可视化对比

本文方法能够实现高分类准确率，关键在于生成的伪图像与真实图像更为相似。简单来说，当飞机伪图像视觉效果真实，则该类型伪图像的深层特征也应该接近真实SSS特征。为了更好地验证本方法的优势，本文采用t-SNE[5]对生成图像进行可视化分析。本文根据表2和表3的分类结果，本文选择了AdaIN生成伪图像、PhotoWCT生成伪图像、本文方法生成伪图像以及真实SSS图像数据集进行分析，如图13所示。其中粉色、蓝色、绿色分别表示飞机、船体、其他类别的伪样本特征，黄色和紫色表示真实SSS飞机和沉船的特征（彩图见期刊电子版）。在图13(a)～13(b)中，可以看出多种伪图像特征是混合的，没有在各自类别的中心位置形成聚集，导致网络无法区分真正的SSS样本。在图13(c)中，SSS飞机的更多特征与飞机伪图像的特征一致，且特征中心也更接近该类别。对于分类网络提取的特征，混合度要低于其他伪图像生成方法。真实SSS目标样本的大部分特征更紧密地聚集在其所属类别中心附近，同时同一类别内的样本特征表现出良好的聚类特性。这使得分类更有效，说明了本文方法的有效性。



(a) AdaIN伪图像的t-SNE图
(a) t-SNE plot of AdaIN pseudo images

(b) PhotoWCT伪图像的t-SNE图
(b) t-SNE plot of PhotoWCT pseudo images

(c) 本文伪图像的t-SNE图
(c) t-SNE plot of this article pseudo images

图 13　伪图像的 t-SNE 可视化对比

Fig. 13　t-SNE visualization comparison of pseudo images

4　结　论

本文提出一种利用伪图像生成的侧扫描声呐图像分类方法。该方法可以有效解决训练样本不足、质量不佳、现有网络模型对训练样本以外的类别目标分类精度较低的问题。其中，本文设计了内容图像特定领域特征削弱模块，削弱光学内容图像与声学图像非关联的像素值、颜色等特定领域特征，促进生成高质量伪SSS图像；IPSTPG网络以光学图像为内容图像，SSS图像为风格图像，通过将光学内容特征与声学风格特征的渲染融合，生成具有内容目标结构和SSS声呐风格的伪图像。为了使生成图像更加真实，本文引入了快速引导增强模块，均衡伪图像中的噪声分布，模拟真实水下噪声而产生的阴影等特征。通过定量和定性的对比实验可知，本算法生成的伪图像质量更高，相较其他方法，本方法的伪图像更适合用于训练SSS图像分类网络。在真实的彩色SSS图像数据集U_r的分类实验中，全局精度和平均精度分别达到了85.03%，77.57%；在真实的灰度SSS图像数据集KLSG的分类实验中，全局精度和平均精度分别达到了86.35%，78.54%。在真实SSS图像的分类实验中，本算法展现出较高的分类精度，进一步验证了算法的先进性和有效性。本文为SSS图像分类任务提供了一种有益的样本扩充方法，具有广泛的应用前景。另外，本文的不足之处在于，当某些SSS图像类别对应的光学图像稀缺或不可获得时，利用已有的光学图像作为内容生成伪SSS图像可能具有一定的局限性，以及分类模型未进行改进以达到更高的分类精度。因此，构建更加多样化的专用光学-SSS图像数据集，缓解SSS图像风格迁移的挑战、提高风格迁移的质量和建立高效的分类模型，以达到更高的分类精度，将是未来的研究方向之一。

作者贡献声明：

徐涛：测量方法的提出，论文构思和撰写；

王朋帅：论文审核与编辑写作；

郭文涛：实施研究和执行调查过程，特别是从事实验研究或收集数据和证据；

蔡磊：提供研究仪器、计算设备资源或其他分析工具；

郑建锋：测量实验的设计及数据整理和分析；

陈亚晨：为研究活动的策划和执行进行管理和协调。

参考文献：

[1] DU X, SUN Y F, SONG Y P, *et al*. Recognition of underwater engineering structures using CNN models and data expansion on side-scan sonar images [J]. *Journal of Marine Science and Engineering*, 2025, 13(3): 424.

[2] PARK J, BAE H S. Effect of seabed type on image segmentation of an underwater object obtained from a side scan sonar using a deep learning approach[J]. *Journal of Marine Science and Engineering*, 2025,

13(2): 242.

[3] 黄海宁, 李宝奇, 刘纪元, 等. 声呐图像水下目标识别综述与展望[J]. 电子与信息学报, 2024, 46(5): 1742-1760.
HUANG H N, LI B Q, LIU J Y, *et al*. Sonar image underwater target recognition: a comprehensive overview and prospects[J]. *Journal of Electronics & Information Technology*, 2024, 46(5): 1742-1760. (in Chinese)

[4] BAI Z Y, XU H L, DING Q C, *et al*. FATL: Frozen-feature augmentation transfer learning for few-shot long-tailed sonar image classification[J]. *Neurocomputing*, 2025, 648: 130652.

[5] JIA Y P, YE X F, LI P, *et al*. Contrastive adaptation on domain augmentation for generalized zero-shot side-scan sonar image classification[J]. *IEEE Transactions on Instrumentation and Measurement*, 2025, 74: 5021013.

[6] BAI Z Y, XU H L, DING Q C, *et al*. Feature contrastive transfer learning for few-shot long-tail sonar image classification[J]. *IEEE Communications Letters*, 2025, 29(3): 562-566.

[7] 袁国铭, 刘海军, 李晓丽, 等. 纹理感知联合颜色直方图特征的水下图像增强[J]. 光学 精密工程, 2024, 32(13): 2112-2127.
YUAN G M, LIU H J, LI X L, *et al*. Underwater image enhancement using joint texture perception and color histogram features[J]. *Opt. Precision Eng.*, 2024, 32(13): 2112-2127. (in Chinese)

[8] 林森, 查子月. 结合色彩补偿与双背景光融合的水下图像复原[J]. 光学 精密工程, 2024, 32(7): 1059-1074.
LIN S, ZHA Z Y. Underwater image restoration method combining color compensation and dual background light fusion[J]. *Opt. Precision Eng.*, 2024, 32(7): 1059-1074. (in Chinese)

[9] MA Q H, JIN S H, BIAN G, *et al*. DBnet: a lightweight dual-backbone target detection model based on side-scan sonar images[J]. *Journal of Marine Science and Engineering*, 2025, 13(1): 155.

[10] GAO X, ZHANG L G, CHEN X Y, *et al*. GCT-YOLOv5: a lightweight and efficient object detection model of real-time side-scan sonar image[J]. *Signal, Image and Video Processing*, 2024, 18(1): 565-574.

[11] XIE B L, ZHANG H M, WANG W H. Side-scan sonar image classification based on joint image deblurring - denoising and pre-trained feature fusion attention network[J]. *Electronics*, 2025, 14(7): 1287.

[12] BAI Z Y, XU H L, DING Q C, *et al*. MCAF-Net: multiscale channel attention fusion network for arbitrary style transfer[J]. *IEEE Transactions on Instrumentation and Measurement*, 2025, 74: 5026813.

[13] XU H L, BAI Z Y, ZHANG X Y, *et al*. MF-SANet: zero-shot side-scan sonar image recognition based on style transfer[J]. *IEEE Geoscience and Remote Sensing Letters*, 2023, 20: 1503105.

[14] LI C L, YE X F, CAO D X, *et al*. Zero shot objects classification method of side scan sonar image based on synthesis of pseudo samples[J]. *Applied Acoustics*, 2021, 173: 107691.

[15] LI Y J, FANG CH, YANG J M, *et al*. Universal style transfer via feature transforms[J]. *arXiv preprint* arXiv: 1705.08086, 2017.

[16] LI Y J, LIU M Y, LI X T, *et al*. A closed-form solution to photorealistic image stylization[C]. *Computer Vision-ECCV* 2018. *Cham*: *Springer International Publishing*, 2018: 468-483.

[17] YAO Y, REN J Q, XIE X S, *et al*. Attention-aware multi-stroke style transfer[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 15-20, 2019, *Long Beach*, *CA*, *USA*. IEEE, 2019: 1467-1475.

[18] DENG Y Y, TANG F, DONG W M, *et al*. Arbitrary video style transfer via multi-channel correlation[J]. *arXiv preprint* arXiv: 2009.08003, 2020.

[19] AN J, HUANG S Y, SONG Y B, *et al*. ArtFlow: unbiased image style transfer via reversible neural flows[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Nashville*, *TN*, *USA*. IEEE, 2021: 862-871.

[20] CHEN H B, ZHAO L, WANG Z Z, *et al*. Artistic style transfer with internal-external learning and contrastive learning[C]. *Neural Information Processing Systems*, 2021

[21] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive in-stance normaliza-

tion[C]. 2017 *IEEE International Conference on Computer Vision* (*ICCV*). *Venice*, *Italy*. IEEE, 2017: 1510-1519.

[22] NATARAJAN P, BASHA K, NAMBIAR A. Synth-sonar: Sonar image synthesis with enhanced diversity and realism via dual diffusion models and gpt prompting [J]. *arXiv preprint* arXiv: 2410.08612, 2024.

[23] LI L, LI Y P, WANG H L, *et al*. Side-scan sonar image generation under zero and few samples for underwater target detection[J]. *Remote Sens*, 2025, 16(22): 4134.

[24] BAI Z Y, XU H L, DING Q C, *et al*. Side-scan sonar image classification with zero-shot and style transfer[J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 5009415.

[25] BAI Z Y, XU H L, ZHANG X Y, *et al*. GC-SANet: arbitrary style transfer with global context self-attentional network[J]. *IEEE Transactions on Multimedia*, 2023, 26: 1407-1420.

[26] DOSOVITSKIY A, BROX T. Generating images with perceptual similarity metrics based on deep networks [J]. *Advances in Neural Information Processing Systems*, 2016: 658-666.

[27] JOHNSON J, ALAHI A, LI F F. *Perceptual Losses for Real-Time Style Transfer and Super-Resolution* [M]. Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 694-711.

[28] PARK D Y, LEE K H. Arbitrary style transfer with style-attentional networks[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 15-20, 2019, *Long Beach*, *CA*, *USA*. IEEE, 2019: 5873-5881.

[29] HE K M, SUN J. Fast guided filter[J]. *arXiv preprint* arXiv: 1505.00996, 2015.

[30] 吴萌, 张倩文, 孙增国, 等. 基于直觉模糊集熵测度和显著特征检测的古铜镜X光图像融合[J]. 光学 精密工程, 2025, 33(2): 262-281.
WU M, ZHANG Q W, SUN Z G, *et al*. X-ray image fusion for ancient bronze mirror based on intuitionistic fuzzy set entropy measure and salient feature detection[J]. *Opt. Precision Eng.*, 2025, 33(2): 262-281. (in Chinese)

[31] LIN T, MAIRE M, BELONGIE S, *et al*. Microsoft COCO: Common objects in context[J]. *arXiv preprint* arXiv: 1405.0312, 2015.

[32] SERGEY K, AARON H, HOLGER W, *et al*. Recognizing image style[J]. *arXiv preprint* arXiv: 1311.3715, 2013.

[33] HE K M, ZHANG X Y, REN S Q, *et al*. Deep residual learning for image recognition[C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 27-30, 2016, *Las Vegas*, *NV*, *USA*. IEEE, 2016: 770-778.

[34] PENG CH Y, JIN SH H, BIAN G, *et al*. SI-GAN: A multi-scale generative adversarial network for underwater sonar image super-resolution [J]. *J. Mar. Sci. Eng*, 2024, 12(7): 1057.

[35] KANG M, ZHANG R, BARNES C, *et al*. Distilling diffusion models into conditional GANs[C]. *Computer Vision-ECCV* 2024. *Cham*: *Springer Nature Switzerland*, 2025: 428-447.

[36] 韩玉兰, 罗轶宏, 崔玉杰, 等. 多模态语义交互的文本图像超分辨率重构[J]. 光学 精密工程, 2025, 33(1): 135-147.
HAN Y L, LUO Y H, CUI Y J, *et al*. Super-resolution reconstruction of text image with multimodal semantic interaction[J]. *Opt. Precision Eng.*, 2025, 33(1): 135-147. (in Chinese)

[37] OKOYE K, HOSSEINI S. *T-test statistics in R: Independent Samples, Paired Sample, and One Sample T-tests*[M]. R Programming. Singapore: Springer Nature Singapore, 2024: 159-186.

[38] PENG Y, LI H P, ZHANG W W, *et al*. Underwater sonar image classification with image disentanglement reconstruction and zero-shot learning [J]. *Remote Sensing*, 2025, 17(1): 134.

[39] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. *arXiv preprint* arXiv: 1409.1556, 2014.

[40] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al*. An image is worth 16x16 words: Transformers for image recognition at scale[J]. *arXiv preprint* arXiv: 2010.11929, 2020.

[41] TAN M X, LE Q. Efficientnetv2: Smaller Models and Faster Training[J]. *arXiv preprint* arXiv: 2104.00298, 2020.

[42] HOWARD A, SANDLER M, CHU G, *et al*.

Searching for MobileNetV3 [J]. *arXiv preprint* arXiv: 1905.02244, 2019.

[43] CHEN X, DUAN Y, HOUTHOOFT R, *et al*. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets [J]. *arXiv preprint* arXiv: 1606.03657, 2016.

[44] HIGGINS I, MATTHEY L, PAL A, *et al*. Beta-VAE: learning basic visual concepts with a constrained variational framework[C]. *5th International Conference on Learning Representations*, 2017.

作者简介：

徐　涛(1981—),男,河南新乡人,博士,副教授,硕士研究生导师,2017年于北京工业大学获得博士学位,主要从水下场景感知、水下自主无人系统研发等方面的研究。E-mail: xutao@hist.edu.cn

王朋帅(2000—),男,河南周口人,硕士研究生,2023年于河南科技学院获得学士学位,现就读于河南科技学院机电学院,主要从事水下场景感知方面的研究。E-mail: wangpengshuai@stu.hist.edu.cn