基于通道分组注意力的无监督图像风格转换模型

doi:10.3969/j.issn.1674-7100.2021.05.011

孙铭一　孙刘杰
李佳昕

上海理工大学
出版印刷与艺术设计学院
上海　200093

摘　要：为了解决无监督图像风格转换模型输出结果的局部伪影和局部特征丢失问题，提出了一种基于通道分组注意力机制的图像风格转换模型。生成器部分采用通道分组注意力残差块，以增强生成器部分对于图像特征的提取以及有效特征的利用；鉴别器部分采用双鉴别器结构，利用增加的局部鉴别器增强对于生成图像细节的鉴别，利用多分辨率尺度的全局鉴别器增强生成图像的内容合理性与结构连贯性。实验结果表明：本模型比起BicycleGAN、MUNIT等模型不但体积更小，而且可以获得更高的NIMA美观度得分以及LPIPS多样性得分；在包装类产品的平面设计迁移应用任务中，本模型同样表现良好。

关键词：无监督；通道注意力机制；图像风格转换
中图分类号：TP751.1　　**文献标志码**：A
文章编号：1674-7100(2021)05-0075-10
引文格式：孙铭一，孙刘杰，李佳昕. 基于通道分组注意力的无监督图像风格转换模型 [J]. 包装学报，2021，13(5)：75-84.

0　引言

图像风格转换是近年来机器视觉领域的研究重点之一。根据图像风格转换模型在训练中是否需要成对的数据，其可分为有监督学习和无监督学习。有监督学习需要成对的数据和人工对数据打标签，导致时间成本过高。无监督学习不需要成对的数据，相较于有监督学习，其数据获取更加简单高效，普适性更高。根据一张输入图像是否能够对应获得多个输出图像，图像风格转换模型可分为单模态模型和多模态模型。在单模态模型中，一张输入图像只能获得一张对应的输出图像，当输入数据不成对时，其局限性便体现在模型的输出结果不确定上。在多模态模型中，一张输入图像可以对应多张输出图像，因而多模态模型能够很好地应对多图像转换任务，例如包装平面设计[1]、文图转换等。

近年来有关多模态无监督图像风格转换的研究越来越多，这些研究都是基于生成对抗网络[2]（generative adversarial networks，GAN）。Choi Y. 等[3]提出了人脸图像多模态转换模型StarGAN，Yu X. M. 等[4]提出了SingleGAN。StarGAN使用星形生成网络结构并在输入中添加目标领域信息，再结合判别器的分类结构和循环重构一致性约束完成图像翻译工作。但是其欠缺对图像重构损失的考虑，因而图像风格转换时某些固定属性会发生变化，导致图像质量下降。SingleGAN则采用类别便签指导方法，在StarGAN的基础之上

收稿日期：2021-05-20
基金项目：上海市科学技术委员会科研计划基金资助项目（18060502500）
作者简介：孙铭一（1997-），女，吉林长春人，上海理工大学硕士生，研究方向为数字图文信息处理，E-mail：1572029707@qq.com
通信作者：孙刘杰（1965-），男，安徽六安人，上海理工大学教授，博士，主要从事印刷机测量与控制技术、数字印前防伪技术、光信息处理技术研究，E-mail：liujiesunx@163.com

对网络结构作出改进。2018 年，Huang X. 等 [5] 提出的 MUNIT 和 Lee H. Y. 等 [6] 提出的 DRIT，均是基于内容与风格分离编码再交叉解码的方法以获得多样的输出。与 DRIT 不同的是，MUNIT 采用自适应实例规范化算法（adaptive instance normalization，AdaIn）的风格特征参数 [7] 来融合内容特征，以实现图像风格转换。DRIT 的转换效果不如 MUNIT，而 MUNIT 的转换效果不如有监督的多模态模型 BicycleGAN[8]。但有监督的多模态模型 BicycleGAN 需要成对的输入数据，这增加了数据集的获取难度，且模型体积庞大。

为了提高无监督模型的输出图像质量，解决局部伪影和局部特征丢失等问题，本课题组在 MUNIT 的基础上，提出了基于通道分组注意力（channel-divided with attention，CDA）的无监督图像风格转换模型。在生成器部分，构建通道分组注意力残差块。在鉴别器部分，利用多分辨率尺度的全局鉴别器对输出图像进行不同分辨率尺度上的鉴别，利用局部鉴别器 [9] 对输出图像局部进行鉴别。

1 无监督图像风格转换模型

1.1 模型结构

本文所提的无监督图像风格转换模型的主要创新点如下：

1）采用通道分组注意力残差块构建生成器。CDA 残差块主要包含通道分组和通道注意力机制（efficient channel attention，ECA）[10-11] 两个模块。通道分组模块能够实现残差块内的跳跃连接，减少特征丢失；ECA 模块能够自适应地调整特征图通道权值，提高网络对有效特征的关注度，并进一步减少模型参数量以及体积。

2）采用双鉴别器结构构建鉴别器。多尺度全局鉴别器对输出图像在多分辨率尺度上进行联级鉴别，以提高输出图像的结构连贯性与内容完整性；局部鉴别器对输入图像进行剪裁，即获得 1/4 的图像，以提高输出图像精度。

3）引入 NIMA（neural image assessment）[12] 美学评分模型评价风格转换图像质量。NIMA 模型对输出图像的真实性进行客观评价，并从图像美学的角度评估图像风格是否美观。将主观评价结果参数化减少了人眼判断的随意性与主观偏差，提高了评价过程的操作便捷性与公平性。

无监督图像风格转换模型的训练网络结构如图 1 所示。测试时，本模型仅需输入单张边缘图像或者实物图像即可得到转换后的风格图像。

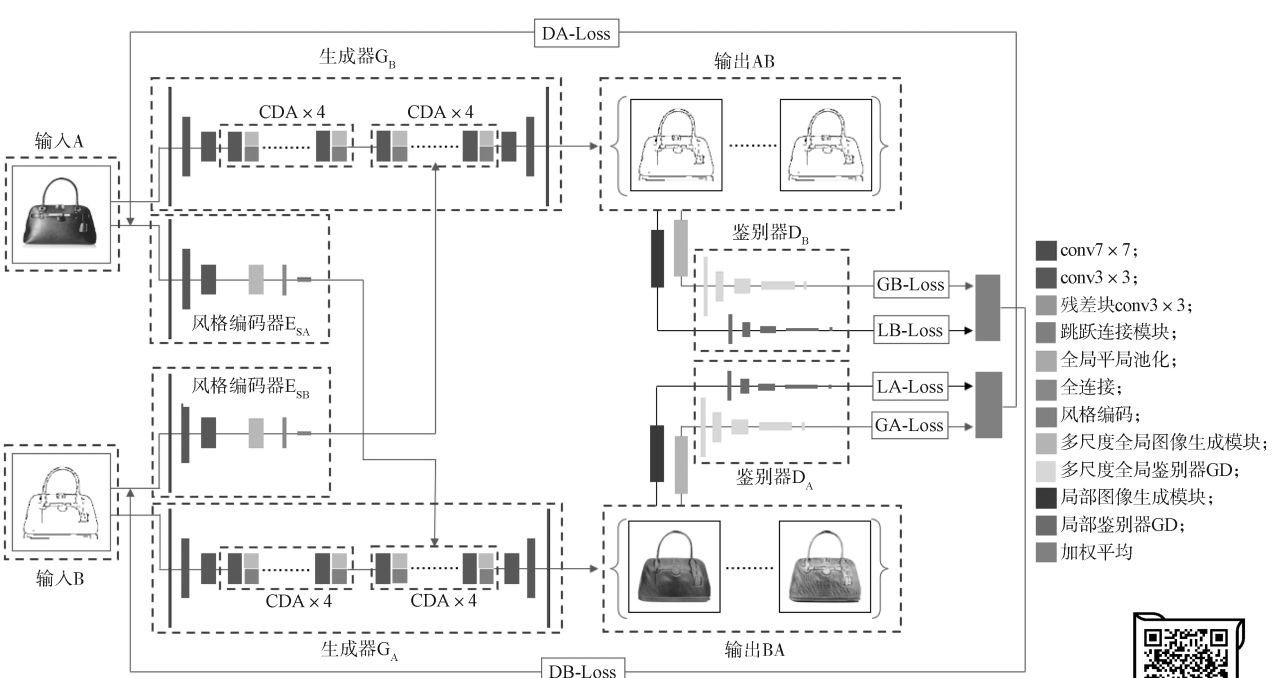


图 1 无监督图像风格转换模型的网络结构

Fig. 1 Network structure of unsupervised model

彩图

在图 1 中，无监督图像风格转换模型为对称结构，包含两个结构相同的生成器 G_A、G_B，两个结构相同的风格编码器 E_{SA}、E_{SB}，以及两个结构相同的鉴别器 D_A、D_B。鉴别器 D_A 由多尺度全局鉴别器 GD 和局部鉴别器 LD 组成。输入图像 A（实物图像）与输入图像 B（边缘图像）互为各自的风格图像，均在生成器和风格编码器完成图像内容和风格的编码，进而实现两个生成器之间的交叉解码，得到的输出图像再进入鉴别器进行鉴别以及前向反馈，最终完成整个训练网络的调整与优化。基于循环一致理论，无监督图像风格转换模型能实现实物图像与边缘图像的互相转换。现有的边缘提取算法，如 Canny 算子、Sobel 算子等，均能检测出清晰完整的边缘，因此本研究不是将图像 A 到图像 AB（边缘图像）的风格转换作为研究重点，而是将图像 B 到图像 BA（实物图像）的风格转换作为研究重点。此转换涉及的主要模块有生成器 G_A、风格编码器 E_{SA}、鉴别器 D_A，转换过程可描述如下：

$$\boldsymbol{BA}=G_{\text{A-decode}}\left(G_{\text{A-encode}}(\boldsymbol{B}),E_{\text{SA}}(\boldsymbol{A})\right)。\tag{1}$$

1.2 风格编码器

风格编码器保留了 MUNIT 中的方法，由两个下采样层、一个池化层以及一个全连接层组成。提取图像的风格编码后，多层感知器（multilayer perceptron，MLP）对其进行加工处理，以一维的 AdaIN 参数形式融合到生成器的解码器中，与内容信息共同解码，从而获得新的风格图像。 MLP 是一种前向的全连接神经网络结构，每一层的单个神经元均与下一层中的所有神经元连接。其中，AdaIN 的工作机制是，给定一个内容图像和一个风格图像，通过调整输入的内容图像的均值和标准差来匹配输入的风格图像，从而实现图像间的风格转换。假设输入的内容图像为 c，输入的风格图像为 s，则 AdaIN 归一化操作公式为

$$\text{AdaIN}(\boldsymbol{c},\boldsymbol{s})=\sigma(\boldsymbol{s})\left(\frac{\mu(\boldsymbol{c})}{\sigma(\boldsymbol{c})}\right)+\mu(\boldsymbol{s})\,,\tag{2}$$

式中：$\mu(\boldsymbol{c})$、$\mu(\boldsymbol{s})$ 分别为内容图像、风格图像的均值；

$\sigma(\boldsymbol{c})$、$\sigma(\boldsymbol{s})$ 分别为内容图像、风格图像的标准差。

式（2）将内容图像与风格图像的均值和标准差对齐，即通过传递特征统计信息在特征空间进行风格转换。以图 1 中的风格编码器 E_{SA} 为例，在图像风格转换过程中，E_{SA} 对输入图像 A 进行编码，提取其风格特征信息 $\boldsymbol{s}_A$，

$$\boldsymbol{s}_A=E_{SA}(\boldsymbol{A})。\tag{3}$$

1.3 生成器

生成器为编码器－解码器结构，其中，编码器由两个下采样层和 4 个通道分组注意力残差块 CDA 组成，解码器由 4 个通道分组注意力残差块 CDA 和两个上采样层组成。编码器和解码器的主要构成模块均为通道分组注意力残差块 CDA，该残差块的构建主要参考了 Gao S. H. 等 [13] 提出的 res2net。res2net 能够以更细的粒度来表示多尺度特征，同时增加了每个块内网络层的感受野，在目标检测、语义分割等机器视觉任务中其有效性得到证实。

边缘图像中，边缘在整张图像中的占比远小于空白部分，过大的感受野会使边缘信息占比更少，导致网络学习很多无效信息，使生成的图像质量下降，出现伪影、空洞等现象。倘若缩小感受野，则会增加参数计算量，加重模型负担。因此，本研究在不改变感受野大小的前提下，将残差块中 basicblock 模块的第二层卷积层按照通道数 n 均分为两个维度相同的网络，一个为常规 3 × 3 卷积，另一个作为浅层信息，通过 concat 跳跃连接 [14] 与卷积后的特征图进行拼接，重构 n 通道特征图。在改进后的残差块末端引入 ECA，构建通道分组注意力残差块结构，如图 2 所示。

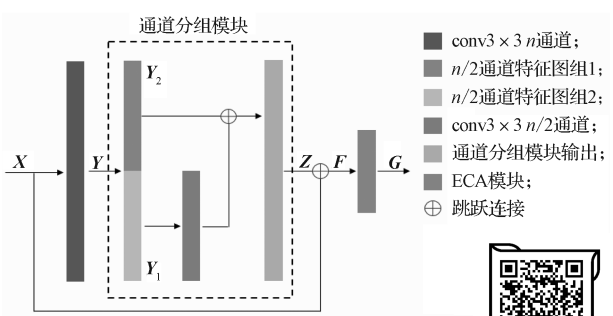


图 2 通道分组注意力残差块结构

Fig. 2 Structure of CDA residual block

彩图

在图 2 中，假设原始输入数据为 $\boldsymbol{X}$，经过第一层卷积后输出 $\boldsymbol{Y}$，

$$\boldsymbol{Y}=\boldsymbol{W}_2\sigma\left(\boldsymbol{W}_1(\boldsymbol{K}_3\otimes\boldsymbol{X})\right),\tag{4}$$

式中：$\boldsymbol{W}_1$、$\boldsymbol{W}_2$ 分别为第一层 3 × 3 卷积、第二层通道分组模块的权重；

σ 为 Relu 激活函数；

$\boldsymbol{K}_3$ 为 3 × 3 卷积核；

⊕为卷积操作。

第二层卷积按照通道数均分为两个网络，分别得到 $n/2$ 通道的特征图 $\boldsymbol{Y}_1$、$\boldsymbol{Y}_2$。$\boldsymbol{Y}_1$ 通过跳跃连接与 $\boldsymbol{Y}_2$ 进行拼接，重构 n 通道特征图 $\boldsymbol{Z}$，

$$\boldsymbol{Y}=\boldsymbol{Y}_1\oplus\boldsymbol{Y}_2, \tag{5}$$

$$\boldsymbol{Z}=\sigma\left(\boldsymbol{W}_{Y_1}(\boldsymbol{K}_3\otimes\boldsymbol{Y}_1)\right)\oplus\boldsymbol{W}_{Y_2}\boldsymbol{Y}_2。 \tag{6}$$

式（5）~（6）中：$\boldsymbol{W}_{Y_1}$、$\boldsymbol{W}_{Y_2}$分别为特征图 $\boldsymbol{Y}_1$、$\boldsymbol{Y}_2$的权重；

⊕为 concat 跳跃连接。

残差块的输出 $\boldsymbol{F}$ 满足如下公式：

$$\boldsymbol{F}=\sigma(\boldsymbol{Z}\oplus\boldsymbol{X})。 \tag{7}$$

通道分组是通过将特征先拆分后融合的策略，使卷积网络能更高效地处理边缘特征。这既实现了边缘特征的深度提取，又实现了浅层边缘特征的重复利用，保留了更多的有效信息[13]。

随后，$\boldsymbol{F}$ 作为 ECA 层的输入数据，进一步完成各通道权重的自适应调整。ECA 是由 Wang Q. L. 等提出，通过不降维且自适应地捕捉图像各特征图之间的跨通道交互，自适应地调整更新各特征图通道的权值，并对各特征图通道间的内部依赖关系进行建模，从而降低模型的复杂程度，提高网络对于有效特征的关注能力。ECA 结构如图 3 所示。

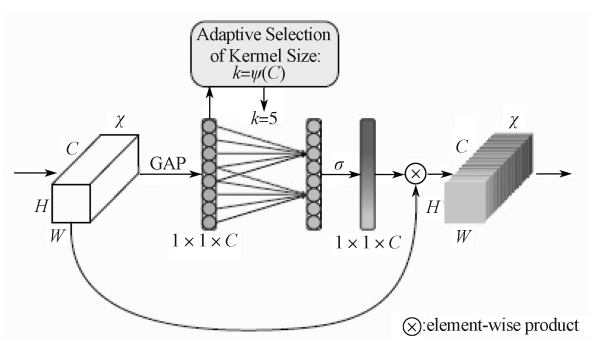


图 3　ECA 结构

Fig. 3　Structure of ECA

ECA 去除了全连接层，在全局平均池化后通过一个可以权重共享的一维卷积对特征图进行学习。该一维卷积涉及超参数 k 即一维卷积的卷积核尺寸，它代表了局部跨通道交互的覆盖率。超参数 k 与输入特征图通道数 C 之间存在如下映射：

$$C=\phi(k)=2^{\gamma\times k-b}, \tag{8}$$

因此，k 为

$$k=\Psi(C)=\left|\frac{\log_2 C}{\gamma}+\frac{b}{\gamma}\right|_{\text{odd}}。 \tag{9}$$

式（8）~（9）中 γ 和 b 为常量，γ=2，b=1。

重构后的特征图 $\boldsymbol{F}$ 可表示为集合 $\{n_i\}(i=1,2,\cdots,n)$，则在 ECA 层中，第 i 个通道的局部跨通道交互权重 w_i 为

$$w_i=\theta\left(\sum_{j=1}^{k}w^j n_i^j\right),\ n_i^j\in\Omega_i^k, \tag{10}$$

式中：θ 为 sigmoid 函数；

w^j 为所有通道间共享的权重；

n_i^j 为集合 Ω_i^k 中的元素，其中 Ω_i^k 为 n_i 的 k 个相邻通道的集合。

可视化通道特征通常表现出一定的局部周期性，因此，通过权重共享的方式捕获局部的跨通道交互，既可以实现有效边缘信息的提取与利用，又避免了捕捉跨所有通道交互所带来的模型复杂度与计算量[10]。经过卷积核大小为 k 的一维卷积后，结合式（7），通道间的信息交互权重矩阵 $\boldsymbol{w}$ 为

$$\boldsymbol{w}=\theta(K_k\otimes\boldsymbol{F})。 \tag{11}$$

式中 K_k 为 $k\times k$ 卷积。

通过权重矩阵 $\boldsymbol{w}$ 实现输出特征图各通道的权值自适应调整，以不降维方式捕获更多有效边缘信息，因此通道分组注意力残差块的最终输出 $\boldsymbol{G}$ 为

$$\boldsymbol{G}=\boldsymbol{w}\times\sigma(\boldsymbol{Z}\oplus\boldsymbol{X})=\boldsymbol{wF}。 \tag{12}$$

1.4　鉴别器

鉴别器由两部分组成，分别是用于鉴别整张输出图像的多分辨率尺度全局鉴别器 GD_A 以及用于鉴别剪裁后局部输出图像的局部鉴别器 LD_A。全局鉴别器 GD_A 采用多尺度结构，即先对图像 BA 进行多次降采样操作，生成大小不同的图像，再通过多分辨率尺度并联鉴别，使输出图像的全局内容连贯和全局结构合理。全局鉴别器 GD_A 包含 3 个不同的鉴别分辨率尺度，分别是 256×256, 128×128, 64×64，每个尺度均输出一个判断值，随后通过加权平均得出整个全局鉴别器的判断值。局部鉴别器 LD_A 为单尺度结构，其输入为剪裁 1/4 的图像 BA 后的图像 BA_{cut}，大小为 128×128。局部鉴别器 LD_A 对局部图像进行鉴别后输出一个判断值，随后与全局鉴别器 GD_A 的判断值通过加权平均方式计算出鉴别器 D_A 的最终判断值。

全局鉴别器和局部鉴别器均采用 PatchGAN[15] 结构。PatchGAN 是将图像分成若干个 70×70 的小块，每个小块输出一个判断值，最终根据得到大小为 $m\times m$ 的矩阵计算判断值。PatchGAN 可以综合考量

整张图像不同部分的影响，使得判断结果更加准确。

在全局鉴别器 GD_A 中，各尺度的鉴别器损失函数 $L^i_{GD_A}$（i=1, 2, 3）为

$$L^i_{GD_A}=E_{A\sim p(\boldsymbol{A})}\left[\log\left(GD^i_A(A_i)\right)\right]+E_{c_B\sim p(c_B),s_A\sim p(s_A)}\left[\log\left(1-GD^i_A\left(G_{A\text{-decode}}(\boldsymbol{c}_B,\boldsymbol{s}_A)\right)\right)\right],\tag{13}$$

式中：E 为期望；

$p(\bullet)$ 为图像的分布；

GD^i_A（i=1, 2, 3）为全局鉴别器中 3 个不同尺度的鉴别器；

$\boldsymbol{c}_B$ 为图像 B 的内容信息；

$\boldsymbol{s}_A$ 为图像 A 的风格信息。

整个全局鉴别器的损失函数 L_{GD_A} 为

$$L_{GD_A}=\sum_{i=1}^{3}\partial_i L^i_{LD_A},\tag{14}$$

式中 ∂_i 为各尺度全局鉴别器的权重系数。

局部鉴别器 LD_A 的损失函数 L_{LD_A} 为

$$L_{LD_A}=E_{A\sim p(\boldsymbol{A})}\left[\log\left(LD_A(\boldsymbol{A}_{cut})\right)\right]+E_{\boldsymbol{c}_B\sim p(c_B),s_A\sim p(s_A)}\left[\log\left(1-LD_A(\boldsymbol{BA}_{cut})\right)\right],\tag{15}$$

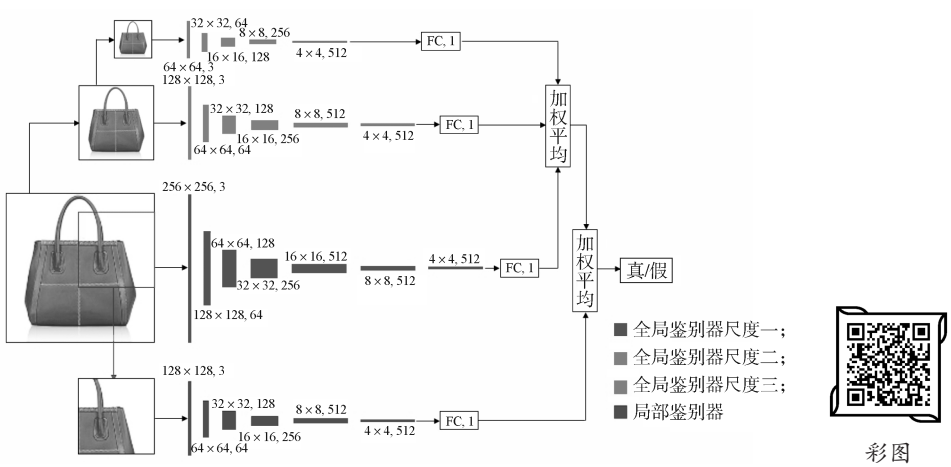
式中 $\boldsymbol{A}_{cut}$ 为剪裁 1/4 的图像 A 后的图像。

鉴别器 D_A 的损失函数 L_{D_A} 为

$$L_{D_A}=\alpha_{GD}\times L_{GD_A}+\beta_{LD}\times L_{LD_A}。\tag{16}$$

式中 α_{CD}、β_{LD} 分别为全局鉴别器、局部鉴别器的权重。

鉴别器的结构如图 4 所示。



彩图

图 4　鉴别器的网络结构

Fig. 4　Network structure of discriminators

1.5　目标函数

无监督图像风格转换模型的总体目标函数为

$$\max_{D_A,D_B}\min_{E_{SA},E_{SB},G_A,G_B}L(E_{SA},E_{SB},G_A,G_B,D_A,D_B)=L_{D_A}+L_{D_B}+\lambda\left(L_{A'\to A}+L_{B'\to B}\right)+\lambda_s\left(L_{s_{A'}\to s_A}+L_{s_{B'}\to s_B}\right)+\lambda_c\left(L_{c_{A'}\to c_A}+L_{c_{B'}\to c_B}\right)。\tag{17}$$

如式（17）所示，目标函数包括 4 个部分。

1）鉴别器损失 L_{D_A} 与 L_{D_B}。

L_{D_A} 为鉴别器 D_A 的损失（见式（15）），L_{D_B} 为鉴别器 D_B 的损失，即

$$L_{D_B}=\alpha_{GD}\times L_{GD_B}+\beta_{LD}\times L_{LD_B}。\tag{18}$$

2）源域图像与重构源域图像间的循环一致性损失 $L_{B'\to B}$ 与 $L_{A'\to A}$。

在生成器 G_A 中，编码器对输入的边缘图像 B 进行编码，提取其内容编码 $\boldsymbol{c}_B$，即

$$\boldsymbol{c}_B=G_{A\text{-encode}}(\boldsymbol{B})。\tag{19}$$

随后，解码器根据内容编码 $\boldsymbol{c}_B$ 以及图像 A 的风格编码 $\boldsymbol{s}_A$，解码获得新的风格图像 BA，即

$$\boldsymbol{BA}=G_{A\text{-decode}}(\boldsymbol{c}_B,\boldsymbol{s}_A)。\tag{20}$$

按照循环一致性理论，生成器获得的风格图像 BA 也能经编码与解码操作得到重构的输入图像 B′。B′的重构过程如下：

$$\boldsymbol{c}_{BA}=G_{A\text{-encode}}(\boldsymbol{BA})；\tag{21}$$

$$\boldsymbol{B}' = G_{\text{B-decode}}(\boldsymbol{c}_{\text{BA}}, \boldsymbol{s}_{\text{B}})。 \quad (22)$$

式（21）~（22）中：$\boldsymbol{c}_{\text{BA}}$ 为风格图像 BA 的内容编码；

$\boldsymbol{s}_{\text{B}}$ 为输入图像 B 的风格编码。

图像 B′与原始输入图像 B 之间的损失为

$$L_{\boldsymbol{B}'\to\boldsymbol{B}} = E_{\boldsymbol{B}\sim p(\boldsymbol{B})}\left[\left\|G_{\text{B-decode}}\left(G_{\text{B-encode}}(\boldsymbol{B}), E_{\text{SB}}(\boldsymbol{B})\right) - \boldsymbol{B}\right\|_1\right]。 \quad (23)$$

同理，图像 A 与重构图像 A′之间的损失为

$$L_{\boldsymbol{A}'\to\boldsymbol{A}} = E_{\boldsymbol{A}\sim p(\boldsymbol{A})}\left[\left\|G_{\text{A-decode}}\left(G_{\text{A-encode}}(\boldsymbol{A}), E_{\text{SA}}(\boldsymbol{A})\right) - \boldsymbol{A}\right\|_1\right]。 \quad (24)$$

3）重构图像的风格编码与原风格编码之间的循环一致性损失 $L_{s_{\text{B}'}\to s_{\text{B}}}$ 与 $L_{s_{\text{A}'}\to s_{\text{A}}}$。

对重构图像 B′进行风格编码，可得风格编码 $\boldsymbol{s}_{\text{B}'}$。$\boldsymbol{s}_{\text{B}'}$ 与 $\boldsymbol{s}_{\text{B}}$（呈正态分布）之间损失应当满足如下约束关系：

$$L_{s_{\text{B}'}\to s_{\text{B}}} = E_{c_{\text{A}}\sim p(c_{\text{A}}), s_{\text{B}}\sim q(s_{\text{B}})}\left[\left\|E_{\text{SB}}\left(G_{\text{B-decode}}(\boldsymbol{c}_{\text{A}}, \boldsymbol{s}_{\text{B}})\right) - \boldsymbol{s}_{\text{B}}\right\|_1\right]。 \quad (25)$$

式中 $q(\cdot)$ 为风格编码 $\boldsymbol{s}_{\text{B}}$ 的分布。

同理，重构图像 A′的风格编码 $\boldsymbol{s}_{\text{A}'}$ 与风格编码 $\boldsymbol{s}_{\text{A}}$ 之间的损失函数为

$$L_{s_{\text{A}'}\to s_{\text{A}}} = E_{c_{\text{B}}\sim p(c_{\text{B}}), s_{\text{A}}\sim q(s_{\text{A}})}\left[\left\|E_{\text{SA}}\left(G_{\text{A-decode}}(\boldsymbol{c}_{\text{B}}, \boldsymbol{s}_{\text{A}})\right) - \boldsymbol{s}_{\text{A}}\right\|_1\right]。 \quad (26)$$

4）重构图像内容编码与原内容编码之间的循环一致性损失 $L_{c_{\text{B}'}\to c_{\text{B}}}$ 与 $L_{c_{\text{A}'}\to c_{\text{A}}}$。

重构图像 B′的内容编码 $\boldsymbol{c}_{\text{B}'}$ 与输入图像 B 的内容编码 $\boldsymbol{c}_{\text{B}}$ 应该是一致的，则 $\boldsymbol{c}_{\text{B}'}$ 与 $\boldsymbol{c}_{\text{B}}$ 之间应当满足如下约束关系：

$$L_{c_{\text{B}'}\to c_{\text{B}}} = E_{c_{\text{B}}\sim p(c_{\text{B}}), s_{\text{A}}\sim q(s_{\text{A}})}\left[\left\|G_{\text{A-encode}}\left(G_{\text{A-decode}}(\boldsymbol{c}_{\text{B}}, \boldsymbol{s}_{\text{A}})\right) - \boldsymbol{c}_{\text{B}}\right\|_1\right]。 \quad (27)$$

重构图像 A′的内容编码 $\boldsymbol{c}_{\text{A}'}$ 与输入图像 A 的内容编码 $\boldsymbol{c}_{\text{A}}$ 之间的损失函数为

$$L_{c_{\text{A}'}\to c_{\text{A}}} = E_{c_{\text{A}}\sim p(c_{\text{A}}), s_{\text{B}}\sim q(s_{\text{B}})}\left[\left\|G_{\text{B-encode}}\left(G_{\text{B-decode}}(\boldsymbol{c}_{\text{A}}, \boldsymbol{s}_{\text{B}})\right) - \boldsymbol{c}_{\text{A}}\right\|_1\right]。 \quad (28)$$

2 实验

本实验是在 Linux18.04 系统、Pytorch1.0 平台完成。训练数据来自 iGAN-project 的手提包图像集。输入图像和输出图像的大小均为 256 × 256。为了验证本文方法的有效性和优越性，用两组不同的测试数据（数据来自 iGAN-project 和网络）测试 BicycleGAN 模型 [8]、MUNIT 模型 [5]、DRIT 模型 [6] 与本模型，利用 NIMA（neural image assessment）距离 [11]、LPIPS（learned perceptual image patch similarity）距离 [16] 评价 4 种模型的输出图像质量，并比较模型的体积和参数量。MUNIT 模型、DRIT 模型和本模型均为无监督模型，输入数据为边缘图像；BicycleGAN 模型为有监督模型，输入数据为一一对应的“边缘 + 实物”图像对。

2.1 评价指标

1）NIMA 距离

引入 NIMA 模型对 4 种模型的输出结果进行真实性评价。NIMA 模型是由谷歌于 2017 年提出的模拟人眼对图片美观度进行打分的模型，通过计算归一化的 EMD（Earth mover’s distance）距离（见式（29））对任意图像生成评分直方图，即给图像进行 1~10 的预测评分。预测评分越高，代表图像质量越高，图像更加美观。

$$EMD(p, p) = \left(\frac{1}{N}\sum_{k=1}^{N}\left|CDF_p(k) - CDF_{\hat{p}}(k)\right|^r\right)^{\frac{1}{r}}, \quad (29)$$

式中：$CDF_p(k)$ 为预测评分的概率累加值，而不是独立的预测获得每一个评分的概率；

$CDF_{\hat{p}}(k)$ 为实际评分的概率累加值。

当标签中的评分越高，则累计概率越大。相比于人眼打分机制，NIMA 模型可以避免人眼主观性较高、观测环境不统一、人眼样本属性不一致等因素带来的偏差。

2）LPIPS 距离

参考 BicycleGAN、MUNIT，引入 LPIPS 距离对 4 种模型的输出结果进行多样性评价。LPIPS 距离由图像深度特征间的 L2 距离加权获得。参考图像 x 与失真图像 x' 之间的距离为

$$d(x, x') = \sum_{l\in L}\frac{1}{H_l W_l}\sum_{h\in H_l, w\in W_l}\left\|\boldsymbol{w}_l \odot \left(\hat{\boldsymbol{y}}^l_{hw} - \hat{\boldsymbol{y}}^l_{0hw}\right)\right\|_2^2, \quad (30)$$

式中：$\hat{\boldsymbol{y}}^l_{hw}$ 为 x 的特征叠加向量；

$\hat{\boldsymbol{y}}^l_{0hw}$ 为 x' 的特征叠加向量；

⊙为矢量 w_l 对通道进行缩放操作。

2.2 iGAN-project 的手提包图像测试实验

从 NIMA 距离、LPIPS 距离以及模型的体积和参数量方面，比较 BicycleGAN 模型、MUNIT 模型、DRIT 模型与本模型的优劣。测试数据来自 iGAN-project 的手提包图像。

训练本模型时，输入图像为手提包边缘图像 B，图像 B 经编码器编码内容信息后与实物图像 A 的风格信息 [17-18] 共同解码获得输出图像 BA，即不同风格的着色手提包图像。部分实验结果如图 5 所示。

图 5 手提包风格转换结果

Fig. 5 Handbags' style conversion results

彩图

与 BicycleGAN、MUNIT 相似，测试实验选取 50 张输入图像，每张输入图像随机采样获得 10 张输出图像，计算 500 张输出图像的 NIMA 距离均值。对 50 张输入图像的每张图像随机采样获得 38 张输出图像，计算 1900 张输出图像的 LPIPS 距离均值。4 种模型的实验结果评价如表 1 所示。

表 1 实验结果评价

Table 1 Evaluation of experimental results

名称	NIMA 距离	LPIPS 距离	体积 /MB	参数量 /M
本模型	4.6784	0.289 ± 0.009	87.8	40.04
BicycleGAN 模型	4.6524	0.261 ± 0.008	219.2	54.79
MUNIT 模型	4.6371	0.287 ± 0.010	120.3	46.60
DRIT 模型	4.6288	0.287 ± 0.008	744.0	65.04

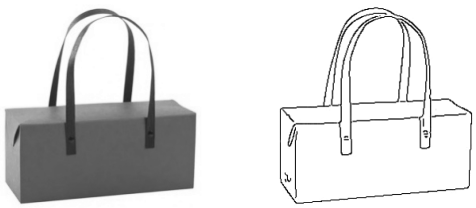
由表 1 可知：从输出图像的美观度来说，本文方法的 NIMA 值最高；从输出结果的多样性来说，本文方法与 MUNIT 和 DRIT 模型差不多，与 BicycleGAN 模型相比，本文方法的多样性提升了约 10%；从模型的体积与参数量来说，本文方法的模型体积最小，参数量最少。可见本文方法能够以更小的模型体积、更少的模型参数量获得更加美观且多样性的输出结果。

2.3 纸质手提盒图像测试实验

在包装的外观设计中，从设计草图到设计稿的过

程，同样可以看作是一次图像间的风格转换。本实验从网络选取 1 张纸质手提盒图像，生成测试数据集。按照实验要求，先将图像统一裁剪为 256 × 256，再利用 Canny 算子提取图像边缘，通过反相操作，获得白色背景的边缘图像（见图 6）。4 种模型的实验结果如图 7 所示。

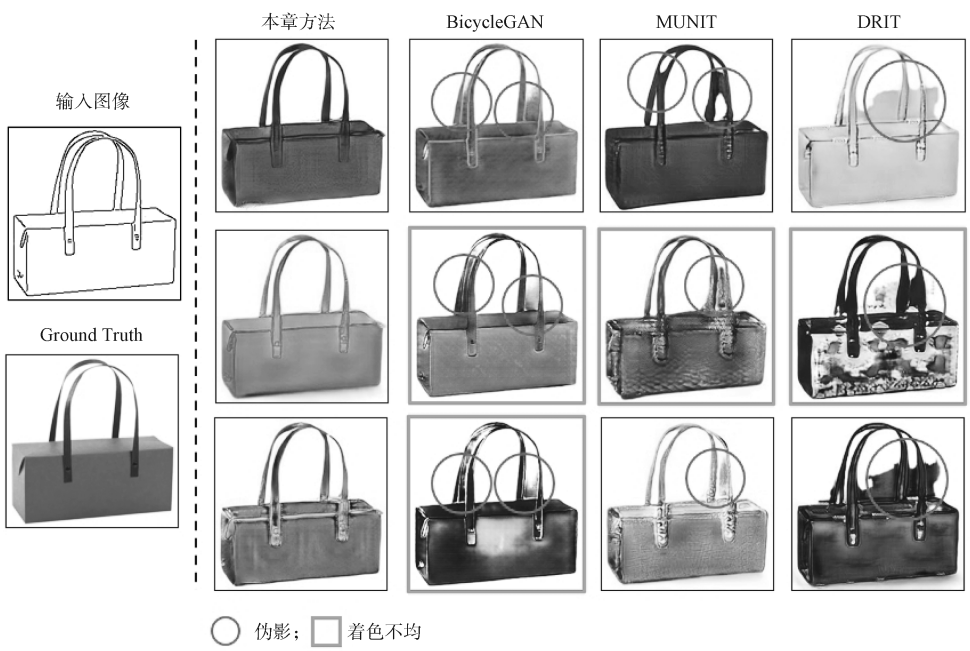
由图 7 可知，BicycleGAN 模型、MUNIT 模型以及 DRIT 模型的输出图像在局部细节上如提手与纸盒的黏合处，均存在严重的伪影，导致输出图像的局部边缘细节不清晰；BicycleGAN 模型和 DRIT 模型的输出结果还出现了着色不均现象，一定程度上影响了输出图像的美观度；本模型的输出图像相对拥有更为清晰的局部细节，在提手与纸盒的黏合处并未出现肉眼可见的大面积伪影，并且图像着色均匀，美观度更高，整体观感更佳。



a）实物图像　　b）边缘图像

图 6　纸质手提盒实物图与边缘图

Fig. 6　Physical image and edge image of paper portable box



彩图

图 7　纸质手提盒可视化结果对比

Fig. 7　Comparison of visualization results of paper portable box

对图 7 中的输出结果计算 NIMA 距离均值，结果如表 2 所示。

表 2　4 种模型的 NIMA 距离

Table 2　NIMA distance of 4 models

指标	本模型	BicycleGAN 模型	MUNIT 模型	DRIT 模型
NIMA 距离	4.5117	3.9648	3.8345	3.7253

由表 2 可知，与 BicycleGAN 模型、MUNIT 模型以及 DRIT 模型相比，本模型在纸质手提盒的平面设计中表现最优，与图 7 的可视化效果相吻合。

上述两组实验结果证明了本文所提的无监督图像风格转换模型在包装产品平面设计迁移应用中的有效性。相较于 icycleGAN 模型、MUNIT 模型以及 DRIT 模型，本模型不仅具有多样性的输出，而且能捕获有效的图像特征，增强图像局部细节。

3　结论

针对多模态无监督图像风格转换模型 MUNIT 模型的输出图像真实性不高的问题，本文提出了一种基于通道分组注意力残差块的双鉴别器无监督模型。首先，在生成器采用基于通道注意力的深度特征提取残差块 CDA，CDA 是编码器与解码器的重要组成模块。CDA 利用跳跃连接提高生成器部分对于浅层图像信

息的提取与利用，并通过 ECA 实现残差块通道权值的自适应调整，进一步提高网络对有效特征信息的关注度。其次，采用并联的多分辨率尺度全局鉴别器与局部鉴别器，重构相应的损失函数。局部鉴别器使生成图像拥有清晰的局部细节，多分辨率尺度全局鉴别器提高生成图像的全局内容连贯性与结构合理性，以更好地实现网络优化，获得更高质量的输出图像。实验结果表明：本模型不仅拥有更小的模型体积，更少的参数量，且在输出图像的 NIMA 美观度评价以及 LPIPS 多样性评价中均取得了更高的得分。此外，在包装类产品的平面设计迁移任务中，本模型也获得了较高的 NIMA 美观度得分，与 BicycleGAN 模型、MUNIT 模型以及 DRIT 模型相比，本模型能够获得局部细节更加清晰、完整的输出图像，减少了伪影、特征丢失等问题的产生，进一步证明了本模型在图像特征提取以及利用等方面的优越性，同时证明了将多模态无监督图像风格转换模型应用于包装设计是可行的，多模态的输出能够为设计工作提供更多的设计思路。

在包装类产品的平面设计中，尽管本模型相较于 BicycleGAN 模型和 MUNIT 模型，在输出图像质量上有一定的提高，但是输出图像还存有小面积的边界模糊以及轻微的伪影，这将是后续研究需要解决的问题。此外，不同包装类型产品相关数据的获取，以及是否需要添加额外的特定约束条件来获得更加真实有效的输出等，也是后续研究方向。

参考文献：

[1] 李佳昕，孙刘杰，王文举 . 基于双鉴别器 GAN 的包装类产品外观设计法 [J]. 包装学报，2020，12(2)：77−83.
LI Jiaxin，SUN Liujie，WANG Wenju. Appearance Design Method of Packaging Product Based on Dual Discriminator GAN[J]. Packaging Journal，2020，12 (2)：77-83.

[2] GOODFELLOW I J，POUGET-ABADIE J，MIRZA M，et al. Generative Adversarial Nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge：MIT，2014，2：2672−2680.

[3] CHOI Y，CHOI M，KIM M，et al. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018：8789−8797.

[4] YU X M，CAI X，YING Z Q，et al. SingleGAN: Image-to-Image Translation by a Single-Generator Network Using Multiple Generative Adversarial Learning[C]//14th Asian Conference on Computer Vision. Perth：[s. n.]，2018，11365：341−356.

[5] HUANG X，LIU M Y，BELONGIE S，et al. Multimodal Unsupervised Image-to-Image Translation[C]//European Conference on Computer Vision. [S. l.]：Springer，2018，11207：179−196.

[6] LEE H Y，TSENG H Y，HUANG J B，et al. Diverse Image-to-Image Translation via Disentangled Representations[C]//European Conference on Computer Vision. [S. l.]：Springer，2018，11207：36−52.

[7] HUANG X，BELONGIE S. Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization[C]//2017 IEEE International Conference on Computer Vision. Venice：IEEE，2017：1510−1519.

[8] ZHU J Y，ZHANG R，PATHAK D，et al. Toward Multimodal Image-to-Image Translation[C]//Advances in Neural Information Processing Systems (NIPS). New York：Curran Associates Inc.，2017：465−476.

[9] IIZUKA S，SIMO-SERRA E，ISHIKAWA H. Globally and Locally Consistent Image Completion[J]. ACM Transactions on Graphics，2017，36(4)：1−14.

[10] WANG Q L，WU B G，ZHU P F，et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle：IEEE，2020：11531−11539.

[11] 张昌凡，孟德志，王燕囡 . 基于轻量化 YOLOv4 的黏稠食品灌装成品缺陷检测 [J]. 包装学报，2021，13(2)：37−45.
ZHANG Changfan，MENG Dezhi，WANG Yannan. Defect Detection in Filling Products of Viscous Food Based on Lightweight YOLOv4[J]. Packaging Journal，2021，13(2)：37−45.

[12] TALEBI H，MILANFAR P. NIMA：Neural Image Assessment[J]. IEEE Transactions on Image Processing，2018，27(8)：3998−4011.

[13] GAO S H，CHENG M M，ZHAO K，et al. Res2Net: A New Multi-Scale Backbone Architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2021，43(2)：652−662.

[14] RONNEBERGER O，FISCHER P，BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]//Medical Image Computing and

Computer-Assisted Intervention（MICCAI 2015），Cham：Spinger，2015：234-241.

[15] ISOLA P，ZHU J Y，ZHOU T H，et al. Image-to-Image Translation with Conditional Adversarial Networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu：IEEE，2017：5967-5976.

[16] ZHANG R，ISOLA P，EFROS A A，et al. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018：586-595.

[17] 黄　慧，史诗琪，郑佳琦 . 以艺术符号视角探析吉祥纹样在文创中的应用 [J]. 家具与室内装饰，2020(8)：83-85.
HUANG Hui，SHI Shiqi，ZHENG Jiaqi. An Analysis of the Application of Auspicious Patterns in Cultural and Creative Design from the Perspective of Artistic Symbols[J]. Furniture & Interior Design，2020(8)：83-85.

[18] 徐沛文，金旭明 . 壮锦纹样在现代家具设计中的应用研究 [J]. 家具与室内装饰，2019(1)：42-43.
XU Peiwen，JIN Xuming. Study on the Application of Zhuang Brocade Pattern in Modern Furniture Design[J]. Furniture & Interior Design，2019(1)：42-43.

（责任编辑：邓　彬）

Unsupervised Image Style Conversion Based on Channel Grouping Attention

SUN Mingyi，SUN Liujie，LI Jiaxin

（College of Communication and Art Design，University of Shanghai for Science and Technology， Shanghai 200093，China）

Abstract：In order to solve the problem of local artifacts and local feature loss in the output of unsupervised image style conversion model, an image style conversion model based on channel grouping attention mechanism was proposed. In the generator part of the model, a channel grouping attention residual block was used to enhance the extraction of image features and the utilization of effective features. In the discriminator part of the model, a dual discriminator structure was adopted, the added local discriminator was used to enhance the identification of the generated image details, and the multi-resolution global discriminator was used to enhance the content rationality and structural coherence of the generated image. The experimental results show that the unsupervised model not only has smaller volume, but also can obtain higher NIMA aesthetic score and LPIPS diversity score than other methods such as BicycleGAN and MUNIT. The model also performs well in the graphic design migration application task of packaging products.

Keywords：unsupervised；efficient channel attention；image style conversion