

# 文献综述：结合生成对抗网络及多角度注意力的图像翻译模型

## 1. 文献信息

- **作者：** 杨百冰, 陈泯融, 叶勇森
- **发表年份：** 2023
- **文献标题：** 结合生成对抗网络及多角度注意力的图像翻译模型
- **期刊/会议名称：** 计算机系统应用
- **DOI/网址：** 10.15888/j.cnki.csa.009059

## 2. 研究背景与目的 (Why)

针对难以在两个差异大的域之间进行图像翻译的问题，本文提出一个新的无监督图像翻译模型，结合了生成对抗网络和多角度注意力，称为 MAGAN。现有注意力机制（如CAM）忽略了通道间特征相关性，导致翻译图像失真。MAGAN通过多角度注意力引导模型关注判别性区域，旨在提升跨域翻译的精度和视觉质量。

相关文献提及：

用于多域图像翻译的 StarGAN v2 只需要映射纹理和颜色样式就可以轻松翻译面部风格 (如面部表情、肤色和性别)。然而，StarGAN v2 在两个差距较大域之间的图像翻译任务上表现一般，一个合理的解释是这些翻译任务中存在复杂的外观变化：即除纹理变换之外，网络还需要夸大图像的局部结构，将局部形状转化为目标域的相应样式。

ContrastGAN[3] 通过调整对象实例的掩码实现对完整图像的注意力语义操作。

UGATIT[5] 将CAM[6] 作为辅助分类器来获取注意力图，引导模型专注于类别判断区域。忽略了像素之间的关联，从而导致生成的图像出现明显的特征失真

## 3. 研究方法与技术路线 (How)

结合生成对抗网络及多角度注意力的图像翻译模型 (MAGAN)：

生成器：由编码器 E、解码器 D 和多角度注意力 MA 组成。设计了一个多角度注意力，由 3 部分组成：全局最大池化 (GMP)、全局平均池化 (GAP) 和空间激活映射 (SAM)，在生成器中取  $p=2$ ，对最具判别力的区域赋予更大的权重。

判别器：由编码器 E、多角度注意力模块 (MA) 和分类器 组成，注意力图关注同一域中真实图像和生成图像的差异， $p$  设置为 1 关注全局和局部信息。

损失函数：包括跨域对抗损失、风格损失、内容重建损失和类别损失，使用预训练的 VGG19 作为感知网络来提取图像的高级语义特征，以提取输入图像和生成图像的深度感知特征

数据集：selfie2anime、cat2dog、horse2zebra 以及vangogh2photo

---

## 4. 核心内容与创新点 (What)

结合生成对抗网络及多角度注意力的图像翻译模型 (MAGAN)，训练生成器来学习具有未配对数据的两个域之间的相互映射。

相同的结构接受不同程度的空间激活（生成器（ $p=2$ ）关注局部判别区域，判别器（ $p=1$ ）平衡全局与局部信息）

结合跨域对抗损失、风格损失、内容重建损失和类别损失，促进网络生成更高质量的图像并使整个网络训练更加稳定

---

## 5. 研究结论与实际意义

实验表明，MAGAN生成了高质量的图像，在风格迁移质量和内容保留上优于现有模型，尤其在差异较大域的翻译任务中表现突出。该模型适用于多种类型的图像翻译任务，如动漫生成、艺术风格迁移等场景，为无监督图像翻译提供了新的技术路径。

---

## 6. 本文的不足与未来展望

注意力模块不能实现内容和风格的完全解耦，所以生成的图像中包含预期外的内容。

在未来的工作中，将研究一种新的模型以完全解耦图像的内容和风格，从而实现更精准的图像翻译。

---

## 7. 个人思考与选题关联

多角度注意力机制的设计

（结合扩散模型：将MA融入扩散过程，提升复杂场景的生成质量。

动态解耦优化：引入可学习参数动态平衡内容与风格，解决当前模型的部分局限性）

---