Image Colorization: A Survey

servuskk
allesgutewh@gmail.com

摘要—在彩色摄影技术普及之前，已经产生了很多黑白图片，它们包含了许多独一无二的记忆。但相比于彩色图片，黑白图片丢失了色彩信息导致视觉效果并不理想。深度学习的兴起为黑白图片上色问题提供了很多解决方案，但由于问题本身的复杂性，AI 上色技术仍未成熟，本文对该领域有代表性的相关工作展开综述，重点关注效果好、热度高、有特色的解决方案。

***Index Terms*—Image colorization, deep learning, colorization review.**

I. 绪论

黑白图上色是一个有趣且有挑战性的任务。彩色图片信息在黑白图中仅保留了一个通道，完成上色需要补充缺失的两个通道的信息，这无疑是非常困难的；同时图片内容涵盖了所有物体类别，且同一物体可以有多种合理的上色方案，让问题的复杂性大大增加 [1]；考虑那些在彩色摄像技术普及之前所拍摄的黑白老照片，还面临分辨率低、图片有损伤等问题。由于问题本身的复杂性，相关研究一直在不断推进，近年来该任务的研究对象从黑白线稿发展到自然界的真实图片，研究方法从数学方法发展到深度学习。

本文将图片上色算法分为用户干预的上色和全自动上色两大类，对近年来有代表性的算法进行简单介绍。

II. 上色算法分类

上色算法研究可以根据颜色的来源分为两大类，其中用户干预的算法又可以分为基于涂鸦和基于范例两类，全自动上色主要使用深度学习方法，可以根据其网络结构划分类别。两个大类特点和问题分析如下。

A. 用户交互上色算法

这类算法又可大致分为基于涂鸦和基于范例两类。其中基于涂鸦的方法需要用户在黑白图像上绘制颜色笔划来指导上色，然后将色彩传播到其它像素，如果两个像素在相似性度量下相似且空间相邻，则分配相同的颜色。涂鸦方法的优点是可以提供较为准确的色彩且鲁棒性强，缺点是需要大量人工。

基于范例的方法则需要与黑白图像相似的彩色图像范例。通过使用范例图像和黑白图像的相似度测量和语意特征关系，将颜色统计信息从彩色图像范例传递到黑白图像。在范例与待上色的图片匹配度较高时可以得到较好的效果，缺点是对范例要求依赖性较高，需要范例图片与黑白图像的相似度较高。

虽然用户干预的算法需要额外的输入，一定程度上增加了输入端的工作量，考虑到自然界中一些物体的颜色例如衣服、建筑物、工艺品等的合理颜色并不是唯一的，用户通过干预能够获得更加符合预期的颜色组合。

B. 完全自动上色算法

该类算法仅需要输入黑白图像无需任何附加信息，能大大减少上色任务的工作量，随着深度学习的发展，成为近期研究的热点。虽然不断有很多效果较好的模型被提出，但由于缺少附加信息的指导，该类算法大多存在色彩不够鲜艳、上色不连贯和伪影等共性问题。

III. 用户交互的上色算法

该部分将介绍有代表性的用户干预上色算法。

A. 基于色彩信息输入的交互上色

线稿上色在本世纪初取得了很多不错的研究成果，其中基于数学的算法无需高算力 GPU 就能提供稳定的线稿上色。Yingge Qu 等在 2006 年提出的 Manga colorization [2] 有效地对包含大量笔画、阴影、半色调和加网的黑白漫画进行着色，由用户进行涂鸦，算法使用 Gabor 小波滤波器获得基于统计的局部图案特征来测量图案连续性，然后通过监视模式连续性的水平集方法传播边界，根据这些信息对线稿精准分割并上色。Daniel Sýkora 等在 2009 年提出的 LazyBrush [3] 兼容多种类型的手绘卡通作品上色，基于图论的 segmentation 方

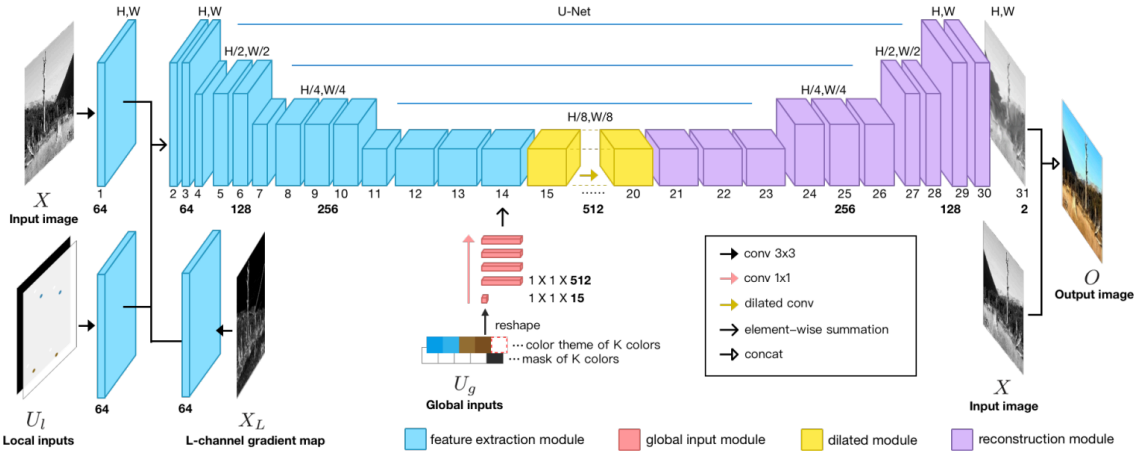

图 1. 整体、局部交互网络结构图

法把上色问题建模为一个“最优化问题”，定义一个“能量函数”来形式化地定义“线稿上色”，当能量函数的值达到最小时，就实现了最优的上色效果。

随着人工智能的兴起，许多研究者将神经网络应用到线稿上色问题。例如 2017 年 Patsorn Sangkloy 等人提出的 Scribbler，[4] 利用一个基于残差连接的 encoder-decoder 网络结构完成涂鸦线稿上色；Lvmin Zhang 等人提出了基于增强残差 Unet 和 GAN 的风格迁移算法，这是一种基于范例的上色算法，其生成器网络结构如图 2所示，输入为一张黑白线稿和一张彩色参考图，彩色参考图经过一个 VGG16 或 VGG19 网络提取特征，黑白线稿经过一个 encoder 结构后，有三个 decoder 与之对应，分别用于直接解码、解码添加了彩色特征的图片和获得最终的彩色图片输出。

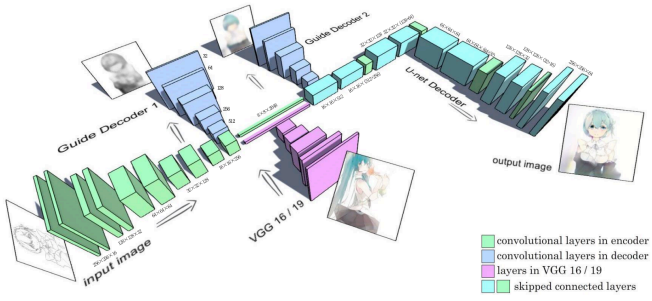

图 2. Scribbler 生成器网络结构图

自然图片相比于线稿包含的内容更加复杂多变，色彩连续性更强，计算复杂度也相应大幅提高，深度学习的产生为解决这类问题提供了可能。在实际操作中，对于自然图片，通常会把 RGB 图片转换为 LAB，因为黑白的灰度图能够提供 L 通道的信息，仅需补充 AB 通道的信息。在用户干预的上色算法中，考虑到需要输入附加信息，附加信息的种类可以是局部的涂鸦也可以是整体风格，Yi Xiao 等在 2019 年提出了一种同时支持整体参考信息输入和局部参考信息输入的深度上色网络 [5]，网络结构如图 1所示，整体采用 Unet 结构，包括蓝色的特征提取网络，粉色的全局色彩输入网络，黄色的细节网络和紫色的特征重建网络。其中特征提取网络的输入由待处理的黑白图片、用户输入的局部信息和黑白图片 L 通道的梯度图三部分构成，输出为三部分结合提取到的特征，特征结合经全局色彩输入网络处理的颜色参考进入细节网络进一步处理后送入重建网络获得最终的彩色图片。这一方法的亮点在于既支持涂鸦参考信息也支持全局色彩范例，即用户既可以指定某一细节的颜色也可以指定整体颜色风格。

B. 基于其他信息输入的交互上色

除了涂鸦和参考图信息作为输入，Jianbo Chen 等人在 2018 年提出的 LBIE 算法 [6] 能够实现使用文字输入指导线稿和自然图片上色；Eungyeup Kim 等人在 2021 年提出的 Deep Edge-Aware Interactive Colorization [7] 先对图片进行自动上色，通过用户反馈并画出有色彩溢出的边缘从而对上色效果进行校正，有效弥补了自动上色色彩溢出的通病。

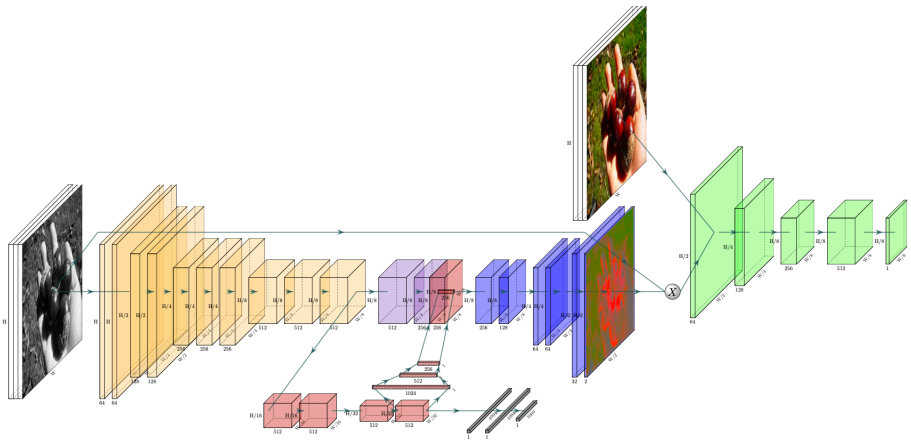

图 3. ChromaGAN 网络结构图

IV. 完全自动上色算法

该部分将介绍有代表性的基于深度学习的自动上色算法网络细节。

A. 备受欢迎的 Encoder-decoder 结构

encoder-decoder 结构多次出现在该领域的论文中，下面介绍三种基于 encoder-decoder 结构的算法。三种算法的效果在其论文发表时均达到了 SOTA，且其网络结构和损失函数也有一定的相似之处。

Patricia Vitoria 等人在 2020 年提出了 ChromaGAN [8]，将对抗性学习方法与语义信息相结合，提出了将色彩、感知信息与语义类别分布结合在一起的三项损失函数。其网络结构如图 3所示，左侧是生成器，分为上下两部分，上面的部分是一个 encoder-decoder 结构，其中黄色的 encoder 部分是 VGG16 [13] 去掉最后三个全连接层构成的，使用预训练的 VGG16 初始化，粉色部分是两个 Conv-BatchNorm-ReLu 模块；下面部分用于输出类别分布向量，红色部分由四个 Conv-BatchNorm-ReLu 和三个全连接层构成，灰色部分输出类分布矢量，用 softmax 函数生成 m 个语义类的概率分布特征；紫色的 decoder 将两个分支合并，用 Conv-ReLu 形式的六层网络加两次上采样处理数据得到输出图像的色度信息。右侧绿色部分是基于 PatchGAN [9] 的鉴别器，关注局部图像块，跟踪所生成图像的高频结构。该算法的亮点在于将语义信息和对抗生存网络结合，取得了比较好的效果。

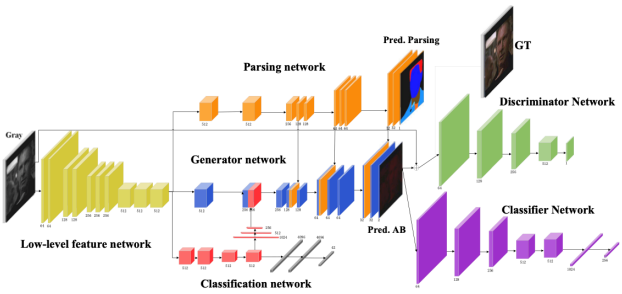

图 4. Focusing on Persons 网络结构图

其后有 Xin Jin 等人的 Focusing on Persons [10] 和 Yuzhi Zhao 等人的 SCGAN [11] 都采用了类似的 endocer-decoder 结构并加入了类别分析，二者的网络结构图如图 4、5所示。Focusing on Persons 在 chromagan 的基础上加入了上方的 Parsing network（橙色部分）来提取人像的特征，同时增加了一个基于 infoGAN [12] 的类别鉴别器（紫色部分）来提高网络对类别的敏感程度。

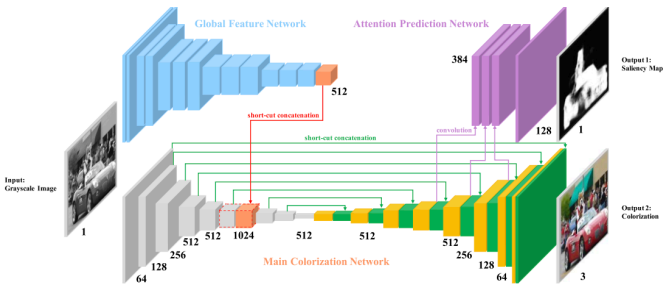

图 5. SCGAN 网络结构图

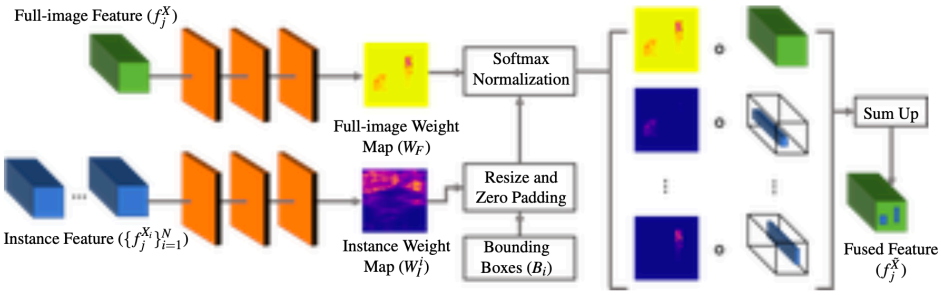

图 6. Instance-aware 网络结构图

SCGAN 的上色主网络也是一个 encoder-decoder 结构，上方的 Global Feature Network（蓝色部分）是一个分类网络，与 ChromaGAN 的 encoder 部分类似，都是基于 VGG16 [13] 分类网络调整而来，SCGAN 的特殊之处在于添加了 Attention Prediction Network（紫色部分）对图片的显著性区域的上色做了优化，其鉴别器使用了两个 PatchGAN 结构分别计算整张图片和显著性区域的 loss。

B. 研究思路创新 *Instcolorization*

目前已有的网络存在一些共性问题，例如颜色偏向蓝绿色不够鲜艳、色彩溢出等，前面提到的几种算法都是用了对抗生成网络来使图片色彩尽可能鲜艳、自然，而解决这些问题除了关注上色任务本身，也有研究提出了一些解决问题的新思路，例如 Jheng-Wei Su 等在 2020 年提出的 Instance-aware Image Colorization6，网络结构如图 6所示。

该算法为解决上色后图片整体颜色趋于一致性的问题，将目标检测、分割加入到上色网络中，将整个上色网络分为整体上色和局部上色两部分，最后训练了一个融合模型将各部分与整体上色的结果融合到一起。在目标检测分割方面，作者采用了 Facebook 团队开发的工具 detectron2 对输入的黑白图像进行目标检测、实例分割，并将分割得到的各个实例作为局部上色网络的输入；在上色方面作者使用了由 Richard Zhang 等人提出的 [15] 当时的 SOTA 模型。在算法效果方面，对于比较明显的多物体图片，该模型很好的区分了不同物体的颜色风格，输出的彩色图片颜色相比其他算法更加鲜艳、有区分度，但对于不能明确划分为多个部分的图片，目标检测不能提供有效信息，此时的输出几乎仍然是黑白的。该算法上色成功与否很大程度依赖目标检测的准确程度，同时由于 detectron2 切割出来的目标是矩形图像块，边缘色彩溢出的问题仍然严重。虽然该方法存在很多问题，但他改变了研究思路，充分利用了开源工具和前人的优秀工作成果，对后续研究有很好的启发性。

C. 巧妙的问题转化

针对自动上色缺失信息过多的问题，Yanze Wu 等人 [16] 在 2021 年提出了一种非常有特点的算法，利用生成的先验色彩指导后续的上色过程，虽然也用到了 encoder-decoder 结构，但整体思路大不相同。

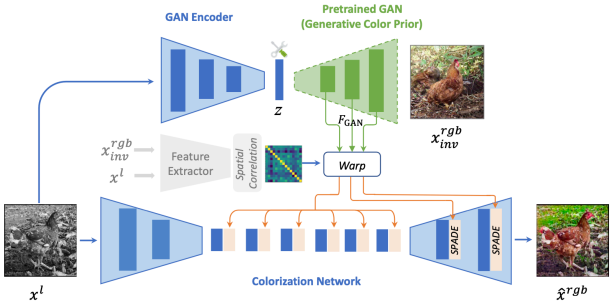

图 7. 先验色彩指导上色网络结构图

其网络结构如图 7所示，网络包括生成色彩先验网络、上色网络、特征调制三部分。生成先验色彩网络的输入是黑白图片，经过编码器得到 latent code，通过预训练的 BigGAN 模型得到与输入图片最相关的多尺度特征；特征调制过程将生成色彩先验网络种的多尺度特征与上色网络中黑白图片对应的多尺度特征进行空间对齐，使用 SPADE 方式调制上色网络；上色网络是一个普通的编解码器结构。输入黑白图片经过编码、调制、解码的过程得到最终的彩色图片输出。该算法的亮点在于跳出了使用 GAN 生成鲜艳色彩的思路，通过一个封装了丰富色彩信息的先验网络获取到参考图为上色提

供范例，后续过程中将自动上色过程转换为基于范例的上色问题，既获得了大量参考信息，又不增加人工工作量。

D. 优秀的工程实践 DeOldify

最后介绍一个公认的效果较好的用于着色和恢复旧图像及视频的深度学习项目 DeOldify [17]，自 2019 年创立以来就备受关注，截至 2021 年 12 月该项目在 GitHub 上已有 14.4k Star，上述深度学习模型的文章中多篇将模型效果与 DeOldify 作对比，在 Instance-aware Image Colorization 的论文中提到“有趣的是，虽然 DeOldify 并没有在基准着色实验中提供最准确的上色结果，但它的上色效果却被更多用户喜爱”[14]。

DeOldify 采用了一种被称作 NoGAN 的新型的、高效的图像到图像的 GAN 训练方法。细节处理效果更好，渲染也更逼真。不同于常规的 GAN，NoGAN 的生成器经过了常规 loss 的预训练而非随机生成，这使得 NoGAN 花费比常规 GAN 架构更少的训练时间。整个生成器网络架构在 U-Net 上使用 ResNet101 主干，并在其中加入了 Self-attention 机制以增强图片色彩的连续性。该项目已在 GitHub 上开源，虽然其网络结构并不复杂，但由于其训练过程中需要多次人工筛选各阶段的最优模型，对模型训练者的经验要求极高，且训练所使用的 fastai 框架用户较少，并未发现有基于 DeOldify 的相关研究，但用到的 Self-attention 机制可以为后续研究提供方向。

参考文献

[1] S. Anwar, M. Tahir, C. Li, A. Mian, F. S. Khan, and A. W. Muzaffar, “Image colorization: A survey and dataset,” *arXiv preprint arXiv:2008.10774*, 2020.

[2] Y. Qu, T.-T. Wong, and P.-A. Heng, “Manga colorization,” *ACM Transactions on Graphics (SIGGRAPH 2006 issue)*, vol. 25, no. 3, pp. 1214–1220, July 2006.

[3] D. Sỳkora, J. Dingliana, and S. Collins, “Lazybrush: Flexible painting tool for hand-drawn cartoons,” in *Computer Graphics Forum*, vol. 28, no. 2. Wiley Online Library, 2009, pp. 599–608.

[4] P. Sangkloy, J. Lu, C. Fang, F. Yu, and J. Hays, “Scribbler: Controlling deep image synthesis with sketch and color,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5400–5409.

[5] L. Zhang, Y. Ji, X. Lin, and C. Liu, “Style transfer for anime sketches with enhanced residual u-net and auxiliary classifier gan,” in *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*. IEEE, 2017, pp. 506–511.

[6] J. Chen, Y. Shen, J. Gao, J. Liu, and X. Liu, “Language-based image editing with recurrent attentive models,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8721–8729.

[7] E. Kim, S. Lee, J. Park, S. Choi, C. Seo, and J. Choo, “Deep edge-aware interactive colorization against color-bleeding effects,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 14 667–14 676.

[8] P. Vitoria, L. Raad, and C. Ballester, “Chromagan: Adversarial picture colorization with semantic class distribution,” 2019.

[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[10] X. Jin, Z. Li, K. Liu, D. Zou, X. Li, X. Zhu, Z. Zhou, Q. Sun, and Q. Liu, “Focusing on persons: Colorizing old images learning from modern historical movies,” in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 1176–1184.

[11] Y. Zhao, L. M. Po, K. W. Cheung, W. Y. Yu, and Y. Rehman, “Scgan: Saliency map-guided colorization with generative adversarial network,” 2020.

[12] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, “Infogan: Interpretable representation learning by information maximizing generative adversarial nets,” in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 2180–2188.

[13] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[14] J. W. Su, H. K. Chu, and J. B. Huang, “Instance-aware image colorization,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[15] R. Zhang, J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, and A. A. Efros, “Real-time user-guided image colorization with learned deep priors,” *arXiv preprint arXiv:1705.02999*, 2017.

[16] Y. Wu, X. Wang, Y. Li, H. Zhang, X. Zhao, and Y. Shan, “Towards vivid and diverse image colorization with generative color prior,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 14 377–14 386.

[17] J. Antic, “Deoldify,” 2019. [Online]. Available: https://github.com/jantic/DeOldify