

# 基于 GAN 的无人机航拍图像重建

曹 琨, 吴 飞, 钱小瑞, 杨照坤

(上海工程技术大学 电子电气工程学院, 上海 201600)

**摘要:** 传统无人机采集传输过程中所传输的数据量常常造成无人机电池的高消耗。针对此类问题, 文中提出一种融合超分辨重建和灰度图像彩色化的 CsRGAN 模型。通过生成网络对低分辨率的灰度图像进行重建: 先将图片进行分辨率放大, 再进行色彩填充, 然后通过判别器进行图片修正, 最终将图片重建为彩色高清图像。实验结果表明, 在固定区域下, 所提出的模型能够在保证成像质量的同时减少无人机航拍的传输数据量, 提高无人机的电池利用率, 且模型具有较强的鲁棒性。

**关键词:** 无人机;超分辨率;色彩填充;图像重建;生成对抗式网络;固定区域

**中图分类号:** TP391

**文献标识码:** A

## Aerial Image Reconstruction of Drone Based on GAN

CAO Kun, WU Fei, QIAN Xiaorui, YANG Zhaokun

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201600, China)

**Abstract:** The amount of data transmitted during the traditional UAV acquisition and transmission process often resulted in high consumption of the UAV battery. Aiming at solving the problem, a CsRGAN model that combined super-resolution reconstruction and gray-scale image colorization was proposed. The low resolution grayscale image was reconstructed by generating a network: the image was first subjected to resolution amplification, color filling was performed, and then the image was corrected by the discriminator, and finally the image was reconstructed into a color high-definition image. The experimental results showed that under the fixed area, the model could reduce the transmission data of the drone aerial photography and improve the battery utilization of the drone under the condition of ensuring the imaging quality. These results proved the model had strong robustness.

**Keywords:** UAV; super-resolution; image colorization; image reconstruction; GAN; fix area

传统无人机航拍中视频采集及传输的数据量较大, 且受限于无线网络的带宽, 需要对视频进行处理方可实时将其传输到地面端。然而视频分辨率越高, 压缩过程越复杂, 将造成高功率消耗。因此在无人机功率受限的条件下, 如何提高无人机电池利用率从而延长无人机的飞行时间, 是一个急需解决的问题。

传统的图像超分辨重建是基于插值方法<sup>[1-3]</sup>, 如 Han 等人<sup>[2]</sup>提出了一种基于各向异性高斯滤波器的边缘自适应插值重建方法, 该方法在移除噪声和边缘附近的阶梯瑕疵时较有优势; Zhang 和 Wu<sup>[3]</sup>提出了一种窗口化的普通 Kriging 插值重建方法, 该方法能较好地保护图像的边缘。但基于插值的方法不考虑图像的内容, 只是简单地进行像素点之间的计

算, 容易丢失大量细节信息, 从而导致模糊。另一种基于重构的超分辨率方法<sup>[5]</sup>结合了图像的降质模型, 解决了基于插值方法无法引入先验信息的问题。该方法虽然对复杂度低的图像效果较好, 但对纹理结构丰富的图像则效果一般。随后研究人员提出了基于学习的方法, 如 ChaoDong 等人<sup>[8]</sup>提出的 SRCNN。该方法结构简单, 结果相较于其他方法更精确, 但是会损失许多图像的细节信息, 同时数据来源并不多, 并不适用于本文中无人机的应用背景。

灰色图像彩色化是将 RGB 的值重新添加到灰度图像的过程<sup>[9]</sup>。传统的色彩填充基本都是通过数字图像传感器进行对颜色进行插值来完成的。例如, Levin 等<sup>[10]</sup>提出能量方程最优化的颜色扩散方法, 其效率比较低。Yatziv 等<sup>[11]</sup>提出了类似的快速图像

和视频的着色方法,但对输入要求较严苛。之后, Welsh 等<sup>[12]</sup>提出了基于参考彩色图像的颜色传递方法,该方法不仅耗时且不能取得满意的结果。

对于固定区域的无人机航拍区域,针对这种特性,文中提出一种融合超分辨率重建和灰色图像彩色化 CsRGAN (Color super Resolution Generative Adversarial Network) 模型。本文首先在保留图像所有细节信息的基础上,将输入的灰色低分辨率图像在颜色超分辨率生成器中先进行放大;然后,在颜色超分辨率生成器部分,输入的彩色帧与灰度帧分别经过多个编码器后融合,再通过多个解码器解码成高清彩色帧并利用判断网络进行比对;根据损失函数对生产网络经过调整,不断地迭代上述过程;最终,在地面段将其还原重构为彩色超清图像。该方法在不影响最终成像的质量的情况下,从减少像素点数量与通道的角度提高了无人机的电池利用率。

## 1 本文算法

### 1.1 生成器网络

本文模型中的生成器网络先对灰度图像进行超分辨率放大,然后再通过多个编解码器进行颜色填充。具体步骤如下:(1)灰度图像首先通过3个编码器不断地减少图像的大小,目的是为了用一个更小的高维的矩阵来表示这个图片;(2)经过6次残差网络来提取特征,使用残差网络即使能使整个网络增加深度,又可防止因梯度弥散而出现不收敛的情况;再通过解码器,用反卷积的方式,将其还原成 $256 \times 256$ 的矩阵,在保持其维度的情况下放大矩阵。最终用亚像素卷积的方式再次放大,这种方法在保留图像所有细节信息的基础对图像进行放大,从而使图像不会发生失真或者形变的情况。然后超分辨率重建后的灰度图片转换为三通道相同灰度值图片,再通过一系列的编码器不断地减少图像的大小,提高图像的通道数。这样做的目的是为了用一个更小的高维矩阵来表示这个图片,最终将其压缩为一个512 维的像素点来表示整个图像。最后,通过解码器,用反卷积的方式,将这个高维的像素点还原成一副随机的彩色图像,在这个过程中,整个网络并不会对图像的整体结构发生太多变化。生成器的内部结构如图1所示。

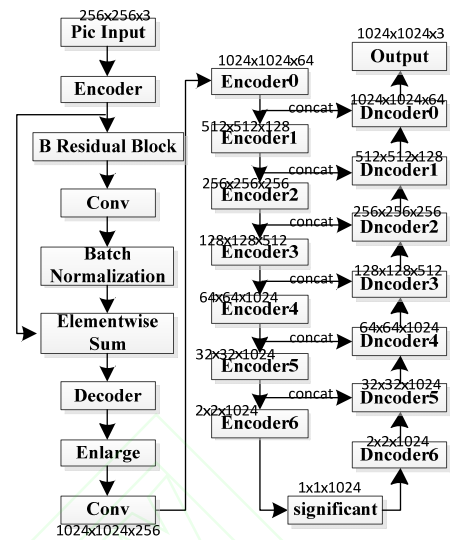


图1 生成器网络内部结构

Figure 1. Internal structure for generator network

但是,单一网络进行卷积反卷积很容易导致网络处于过拟合的状态,特别是当网络有了6层以上卷积时,图像特征包含了大量的高维特征的同时忽略了过多的低层特征信息。因此,本文借鉴 U-net 的思想,使用跳层连接,通过在每一层的输入和输出之间共享大量的低级信息来保证在解码器中反卷积时不会缺少低层的特征信息。

### 1.2 判别器网络

本文模型中的判别器网络分别通过卷积将生成图片和真实图片同时形成一个维度为 $128 \times 256 \times 256$ 大小的矩阵,再将两者链接起来。该方式可避免因将过大尺寸图片送入解码模块,使得最后一层维度过高而出现过拟合的情况。判别器网络内部结构如图2所示。

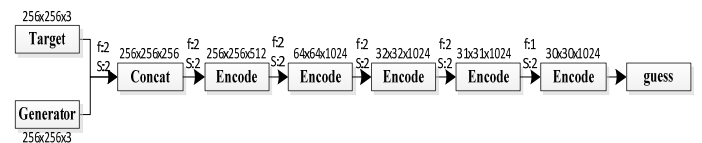


图2 判别器网络内部结构

Figure 2. Internal structure for discriminator network

本文的判别网络借鉴了自编码器的思想,网络的输入层先将真实数据库和生成器生成的图片拼接成一张 $256 \times 256 \times 6$ 的图像。输入拼接的六通道的图像矩阵,经过3个编码块压缩为 $32 \times 32$ 大小的256通道的特征矩阵,编码模块由一个卷积核为 $3 \times 3$ ,步长为1的卷积层、一个批量正则化层(Batch Normalization)和一个修正线性单元激活层组成。

最终输出一个大小为  $30 \times 30$ ，通道为 512 的特征矩阵，其中每个像素值(0~1)表示未知图像对应部分的置信程度。

### 1.3 损失函数

本文使用生成对抗网络来完成图像的重建，所以沿用了 Goodfellow<sup>[14]</sup>所提出的生成对抗网络原型的损失函数，如式 (1) 所示

$$V(D, G) = \min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

本文的生成网络中包括超分辨率重建和灰色图像彩色化两部分，所以本文引用超分辨率重建部分损失  $G_1$  和灰图像彩色化损失  $G_2$

$$G_g = G_1 + G_2 \quad (2)$$

超分辨率重建部分整体损失函数如式 (3) 和 (4) 所示。针对于灰度图像的放大部分，考虑到图像的结构相似性，本文采用 SSIM 来代替 MSE，同时仍然使用预先训练完成的 VGG19 作为网络损失函数。

$$G_1 = \arg \min \frac{1}{N} \sum_{n=1}^N l_{CR}(G, T_1) \quad (3)$$

$$l_{CR} = l_{SSIM}^{CR} + l_{VGG19}^{CR} \quad (4)$$

灰色图像彩色化的整体损失如式 (5) 所示，其中  $L_1$  范数的损失函数为式 (6)，同时增加 HSV 的值作为第二部分的损失函数 (7)

$$G_2 = l_{L_1}^{CR} + l_{HSV}^{CR} \quad (5)$$

$$l_{L_1}^{CR} = \text{avg} \|T - G\| \quad (6)$$

$$l_{HSV}^{CR} = \text{avg} \left[ 0.6 \|T_H - G_H\|_{MSE} + 0.2 \|T_S - G_S\| + 0.2 \|T_V - G_V\| \right] \quad (7)$$

本文的总损失函数需要添加  $L_1$  范数和  $L_{HSV}$ ，目的是增加真实图片和合成输出的相匹配度，在本文的数据集中实验结果如表 1 所示。

表 1 不同网络损失

Table 1. Loss of different networks

Loss	Per-pixel acc.	Per-class acc.	Class IOU
$L_1 + \text{GAN}$	0.68	0.19	0.14
$L_1 + \text{GAN} + L_{HSV}$	0.70	0.22	0.15

最终本文的损失函数如式 (8) 所示。

$$G = G^* + \lambda [l_{L_1}(G) + l_{HSV}(G)] \quad (8)$$

其中，

$$l_{L_1}(G) = E_{T, Z \sim p_{data}} [\|T - G(Z)\|_1] \quad (9)$$

$$l_{HSV}(G) = E_{T, Z \sim p_{data}} [\|T_{HSV} - G_{HSV}\|_{MSE}] \quad (10)$$

## 2 实验结果与分析

### 2.1 实验环境与数据集

本文实验环境：Ubuntu16.04 平台上由配有 Tensorflow1.20 框架的 python 编程实现，处理器为 Intel Core i7-6300HQ, 2.9 GHz 16 核 CPU，内存为 64 GB，显卡为 GTX1080Ti，显存为 8 GB，训练共使用了 30 000 张图片，训练时长 144 小时。如图 3 所示，在校园的图书馆到行政楼，无人机按照逆时针方向按照 A B C D A 的顺序进行拍摄。实验数据集是无人机从所有的实验图像是低分辨率和高分辨率之间的比例因子是  $4x$ ，这相当于图像像素减少了  $16x$ 。



图 3 测试区域

Figure 3. The test area

### 2.2 网络性能与实验结果

在本实验中，只考虑 YCrCb 颜色空间的亮度通道，因此本文中第一层/最后一层中  $c=1$ 。两个色光通道是为了显示，而不是为了进行训练和测试。本文的方法可以通过设置  $c=3$  来扩展到对颜色图像的直接训练。但是在本文使用  $c=1$  是为了与以前的方法进行比较。

为了避免在训练过程中产生边界效应，所有的卷积层都没有填充，而网络产生的输出会更小。在处理测试图像中，卷积神经网络可以应用在任意大小的图像上。在测试期间，所有的卷积层都得到了足够的零填充，因此输出图像的大小与输入的大小相同。为了解决边界效应，在每个卷积层中，每个像素的输出（在 ReLU 之前）都是通过有效输入像素的数量来规格化的，这可以预先计算出来。

每一层的滤波器权重由随机抽取的高斯分布随机抽取，零均值和标准差 0.001。学习速率是前两层的  $10^{-4}$ ，最后一层是  $10^{-5}$ 。从实验经验上发现，在最后一层中，一个较小的学习速率对于网络的收敛是很重要的。实验效果图具体如图 4~图 7 所示。



(a)

(b)

(c)



图 4 序列 15

(a) 输入图片 (b) 输出图片 (c) 目标原图

Figure 4. Frame 15

(a)Input picture (b)Output picture (c)Target picture



(a) (b) (c)

图 5 序列 30

(a) 输入图片 (b) 输出图片 (c) 目标原图

Figure 5. Frame 30

(a)Input picture (b)Output picture (c)Target picture



(a) (b) (c)

图 6 序列 81

(a) 输入图片 (b) 输出图片 (c) 目标原图

Figure 6. Frame 81

(a)Input picture (b)Output picture (c)Target picture



(a) (b) (c)

图 7 序列 130

(a) 输入图片 (b) 输出图片 (c) 目标原图

Figure 7. Frame 130

(a)Input picture (b)Output picture (c)Target picture

图中 4 个序列的图像为重建的效果, 基本符合实验预期。接下来对从本模型中输出的图片与原始的目标图片进行比较和分析, 具体如表 2 所示。由表可知, 彩色与黑白图像之间的差距较大, 最大为 35.2 倍, 平均为 33.9 倍。

表 2 视频图像大小

Table 2. Video size

序列	输入图像/kB	输出图像/kB	差值倍数
15	27	875	32.4
30	28	935	33.4
81	39	1 356	34.8
130	30	1 057	35.2

上述视频序列中的客观评价指标见表 3。

表 3 性能指标

Table 3. Performance index

序列	保真性 PSNR	鲁棒性 NC	结构相似性
			SSIM
15	28.617 190	0.997 791	0.981 915
30	28.815 120	0.997 691	0.981 965
81	27.958 641	0.997 781	0.982 935
130	21.758 541	0.997 715	0.981 915

使用本文方法与普通无人机航拍视频传输方法 [15]及论文[16]中基于仿射模型的无人机视频压缩方法, 进行传输 300 帧视频所需时间的方法进行比较, 结果如表 4 所示。

表 4 不同方法用时比较

Table 4. Times of different networks

方法	数据量 /kB·s <sup>-1</sup>	传输视频分 辨率	播放视频分 辨率	300 帧视频 传输时间/s
文献[15]	10 640.228	640×480	640×480	20.66
文献[16]	12 702.890	640×480	640×480	17.31
本文方法	10 640.228	256×256	1024×1024	2.25

由表 2~表 4 得出, 相比于其他方法, 本文提出的传输低分辨率灰度图像在接收端进行超分辨率色彩还原的方法模型, 其传输的图像在视觉上效果与原始效果相差不大, 且传输的数据量更小, 传输时间更短, 模型鲁棒性较强。

### 3 结束语

本文通过融合超分辨率重建和色彩重建所建立的 CsRGAN 模型, 该模型能够重建灰色低分辨率图像, 将其还原为高相似度的彩色高清图像。实验结果表明, 本文提出的模型结构简单且具有较高的重建精度和较强的鲁棒性。未来将在模型中加入语义分析网络, 以便进一步改善重构结果。对于无人机航拍来说, 该模型能够在视频质量约束条件下, 大大减少视频压缩及传输部分的能耗, 大幅延长无人机的续航时间。

### 参考文献

- [1] 管超.基于稀疏表示理论的图像超分辨率重构算法研究[D].上海:上海交通大学,2013.  
Guan Chao.Research of image super-resolution reconstruction based on sparse representation theory[D].Shanghai:Shanghai JiaoTong University, 2013.
- [2] Han J W,Suryanto,Kim J H,et al.New edge-adaptive image interpolation using anisotropic Gaussian filters[J]. Digital Signal Processing,2013,23(1):110-117.
- [3] Zhang Q,Wu J.Image super-resolution using windowed ordinary Kriging interpolation[J].Optics Communications, 2015,336(336):140-145.
- [4] Freeman W T,Jones T R,Pasztor E C.Example-based super-resolution[J].Computer Graphics & Applications IEEE,2002,22(2):56-65.
- [5] Zhang K,Gao X,Tao D,et al.Single image super-resolution with non-local means and steering kernel regression[J].IEEE Transactions on Image Processing,2012,21(11):4544-56.
- [6] Bevilacqua M,Roumy A,Guillemot C,et al.Neighbor embedding based single-image super-resolution using Semi-Nonnegative Matrix Factorization[C].Kyoto:IEEE International Conference on Acoustics,Speech and Signal Processing,2012.
- [7] Kim C,Choi K,Ra J B.Example-based super-resolution via structure analysis of patches[J].IEEE Signal Processing Letters,2013,20(4):407-410.
- [8] Ledig C,Theis L,Huszar F,et al.Photo-realistic single image super-resolution using a generative adversarial network[C].Honolulu:30<sup>th</sup> IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),2017.
- [9] 冈萨雷斯.数字图像处理[M]北京:电子工业出版社,2005.  
Gonzalez.Digital image processing[M].Beijing: Publishing House of Electronics Industry,2005.
- [10] Levin A,Lischinski D,Weiss Y. Colorization using optimization[J].Image and Vision Computing, 2003,46(3):197-202.
- [11] Yatziv L,Bartasaghi A,Sapiro G.O(N) implementation of the fast marching algorithm[J].Journal of Computational Physics,2006,212(2):393-399.
- [12] Welsh T,Ashikhmin M,Mueller K.Transferring color to greyscale images[J].ACM Transactions on Graphics,2002, 21(3):277-280.
- [13] Isola P,Zhu J Y,Zhou T,et al.Image-to-image translation with conditional adversarial networks[C].Honolulu:30<sup>th</sup> IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),2017.
- [14] Goodfellow I J,Pouget-Abadie J,Mirza M,et al. Generative adversarial networks[J].Advances in Neural Information Processing Systems,2014(3):2672-2680.
- [15] Richardson I E G.H.264 and MPEG-4 video compression: video coding for next-generation multimedia[M].Hoboken:Wiley Press,2004.
- [16] 沈秋,李小凡,孔繁铨,等.基于仿射模型的无人机视频实时压缩算法[J].电子与信息学报,2014, 36(12):2855-2860.  
Shen Qiu,Li Xiaofan,Kong Fanqiang,et al.A real-time video compression for UAV based on affine model[J]. Journal of Electronics & Information Technology,2014, 36(12):2855-2860.

收稿日期: 2018-08-13

#### 基金项目

国家自然科学基金(61272097);上海市科技学术委员会资助项目(13510501400);上海市科委重点项目(18511101600) National Natural Science Foundation of China (61272097); Shanghai Municipal Committee of Science and Technology Project(13510501400); Shanghai Municipal Committee of Science and Technology Project(18511101600)

#### 作者简介

曹琨(1994-),女,硕士。研究方向:深度学习、图像处理。吴飞(1968-),男,博士,教授。研究方向:计算机组织与系统结构、分布式多媒体技术。