

基于深度学习的北半球 2.5 米空间分辨率建筑物高度超分辨率估计方法

摘要

摘要：建筑物高度是评价城市发展的重要指标。在现有研究中，大规模建筑物高度估计多基于粗空间分辨率数据（例如 10 米、500 米和 1000 米），这限制了对城市区域内建筑物高度差异的深入分析。相比之下，高分辨率影像（分辨率小于 5 米）虽能支持更精细的建筑物高度估计，但往往受制于空间覆盖范围有限且难以公开获取的局限性。在此背景下，该论文提出了一种基于深度学习的超分辨率方法，旨在利用 Sentinel-1/2 数据生成分辨率达到 2.5 米的建筑物高度图。该方法由两个关键模块组成：其一是超分辨率模块（SR），负责提取高分辨率特征；其二是高度分层估计模块（HS），通过引导网络学习不同高度区间的特征，有效缓解高度值分布不均的问题。为支持相关研究，该论文作者还构建了一个覆盖范围广泛的开放式建筑物高度数据集，包含 45,000 个样本，覆盖北半球多个城市区域，包括了中国、美国本土和欧洲。

关键词：超分辨率；建筑物高度；Sentinel-1/2；深度学习

1 引言

建筑物高度是评价城市发展的重要指标。快速的城市化进程对土地资源的需求日益增长，用于建设住宅、商业和工业设施的空间变得尤为紧张。城市扩张因此在水平和垂直两个维度上同时展开。尤其是在垂直方向，高层建筑不仅能够在有限的土地上容纳迅速增加的人口，还能为多种经济活动提供额外的空间。然而，这类建筑通常伴随着更高的能源消耗。通过合理规划不同高度的建筑布局，可以有效缓解城市热岛效应，同时提升热舒适度 [5]。在此背景下，遥感技术被广泛应用于建筑物高度的估算，不仅在区域尺度上展现出重要价值，也在全球范围内发挥了重要作用。此外，建筑高度数据在多个领域中发挥着关键作用，例如用于模拟空气污染物的扩散路径、进行人口分布估算、绘制物质存量地图以及推动数字孪生城市的构建。

遥感技术已广泛应用于大陆和全球尺度的建筑物高度估算。在以往的研究中，有人用多源遥感数据建立随机森林模型，预测中国、美国和欧洲在空间分辨率为 1km 的建筑物高度。Wu 等 [18] 利用 Sentinel-1/2 数据构建了随机森林模型，以反演中国 10 米分辨率的建筑物高度。与之相比，Huang 等 [10] 也基于 Sentinel-1/2 影像开发了类似的随机森林模型，完成了同样的任务。不同于上述方法，蔡等 [2] 人提出了一种基于深度学习的建筑物高度预测模型，同样使用 10 米分辨率的 Sentinel-1/2 数据，并在覆盖 63 个中国城市的实验中取得了显著效果。在计算机视觉领域，Shi et al. (2016) [15] 提出了一个高效的亚像素卷积神经网络，能够

从低分辨率输入图像中恢复出高分辨率的细节。Lim et al. (2017) [13] 则提出了一种残差超分辨率网络，并在多个任务中展现了优异的表现。与这些方法不同，Wang et al. (2021b) [16] 采用了增强型超分辨率生成对抗网络 (ESRGAN)，在处理现实世界退化图像时取得了显著的效果。进一步的研究验证了基于 GAN 的方法在 Sentinel-2 和高分辨率遥感影像上的有效性 [17]。尽管如此，基于 GAN 的超分辨率技术在建筑物高度估计中的应用仍然是一个相对较少探索的领域。

近年来，随着 GAN 网络在遥感图像有效性的验证。因此作者引入了一个基于深度学习的增强型超分辨率生成对抗网络 (ESRGAN) 方法，使用开放的 Sentinel-1 和 Sentinel-2 图像使用 2.5 米的空间分辨率来估计建筑物高度。

2 相关工作

2.1 图像超分辨率领域

自 SRCNN [7] 以来，图像超分辨率领域 [6] 经历了各种各样的发展。为了获得视觉上令人满意的结果，通常采用生成对抗网络 [8] 作为损失监督，以使解决方案更接近自然流形 [12]。大多数方法假设双三次下采样核，并且通常在真实图像中失败。近几年来还在图像恢复之前结合了强化学习或 GAN [4]。在 Blind SR 领域，已经有一些出色的探索。第一类涉及显式退化表示，通常由两个部分组成：退化预测和条件恢复。上述两个部分可以单独执行 [1]，也可以联合执行 [9]。这些方法依赖于预定义的退化表示（例如，退化类型和级别），并且通常考虑简单的合成退化。此外，不准确的退化估计将不可避免地导致伪影。另一类是获得/生成尽可能接近真实数据的训练对，然后训练统一的网络来解决 Blind SR。训练对通常是 1) 用特定相机捕获，然后进行繁琐的对齐 [3]；2) 或直接从具有循环一致性损失的未配对数据中学习 [14]；3) 或与估计的模糊核和提取的噪声块合成 [11]。然而，1) 捕获的数据仅限于与特定相机相关的退化，因此不能很好地推广到其他真实图像；2) 使用未配对数据学习细粒度的退化具有挑战性，结果通常不令人满意。

3 本文方法

3.1 本文方法概述

该方法由两个主要部分组成：1) 超分辨率 (SR) 模块；2) 高度分层估计 (HS) 模块。首先，通过 SR 模块对 Sentinel-2 影像进行处理，以提取高分辨率特征 (HR features)。接着，在 HS 模块中，我们结合了 Sentinel-1 和 Sentinel-2 数据，用于同时预测建筑物的高度和其所属的高度等级。最后，将得到的 HR features 融合进 HS 模块，以提升建筑物高度和高度分层的空间分辨率，如图 1 所示。

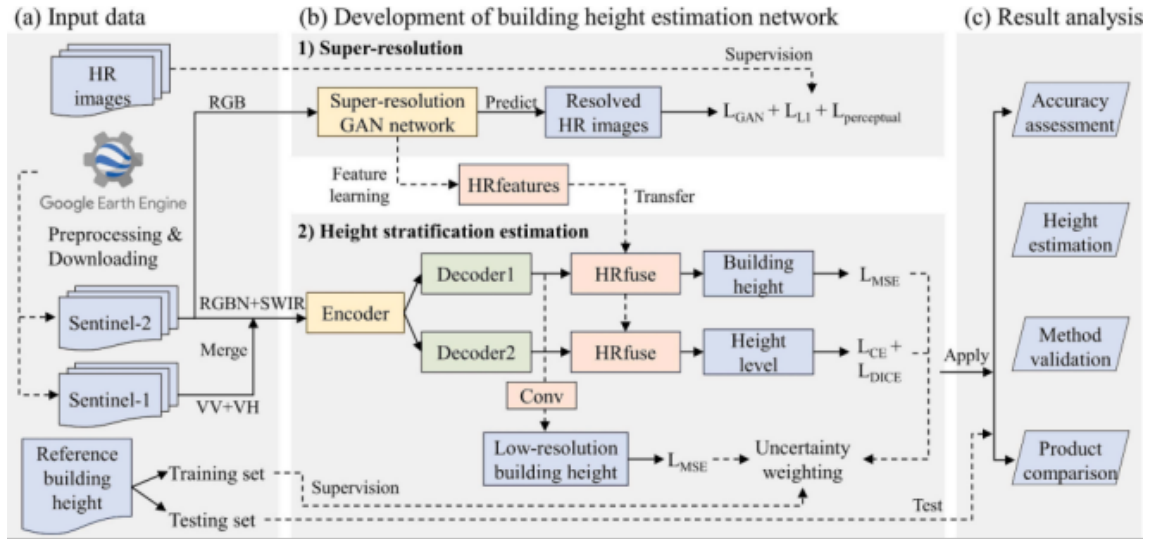


图 1. 方法示意图

3.2 超分辨率模块

该论文的作者引用了 Real-ESRGAN 网络作为超分辨率模块，以提高原始 Sentinel-2 图像的空间分辨率。该网络由生成器和鉴别器组成。生成器主要由残差中的残差密集块 (RRDB) 组成，RRDB 块的数量设置为 23（已经被证明该设置在遥感图像中是有效的），旨在从低分辨率图像中生成高分辨率图像。鉴别器采用带谱归一化的 U-Net 结构（可以提高训练的稳定性），旨在区分参考图像和生成的图像。在该论文中通过使用 Sentinel-2 数据和高分辨率 (HR) 图像对 Real-ESRGAN 网络进行了微调，并从下图 2 中的 Conv3 中提取了高分辨率特征，这些特征可以用于后续的分析。

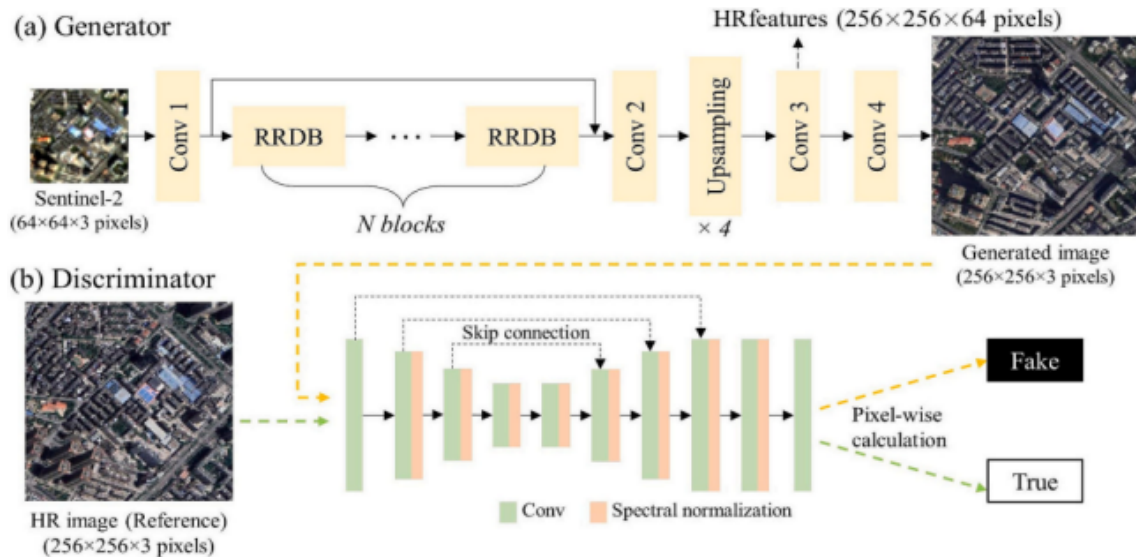


图 2. 超分辨率模块

3.3 高度分层估计模块

该模块设计了一个双分支建筑物高度与分层预测网络，该网络由一个编码器 (Efficient-Net B4) 和两个解码器 (U-Net) 组成。一个解码器用于估计建筑物高度，一个解码器用于高度

分层预测。为解决建筑物高度不平衡问题（非建筑物像素占 80%），将高度分为了 [0,3)、[3,12)、[12,21)、[21,30)、[30, 60)、[60, 90) 和 90 七个区间。并且采用了逆频率加权法，如公式 1，增加高层建筑的权重并降低底层建筑的权重。

$$W = \frac{\sum_{t=1}^7 \sqrt{s_t}}{\sqrt{s_t}} \quad (1)$$

为了提高预测结果的空间分辨率，该论文开发了一个高分辨率融合模块 (HRfuse)，由上采样、残差块和卷积层组成。LRfeatures 为解码器最后一层的低空间分辨率，HRfeatures 为来自超分辨率模块的输出。上采样块包括两个卷积层和两个像素混洗层，其中像素混洗层能够学习复杂的上采样滤波器，从而模拟低分辨率与高分辨率图像之间的映射关系。残差块由三个卷积层、三个批量归一化层以及三个 ReLU 激活层组成，并且已经被证实能够有效地自动提取具有判别能力的特征，如下图 3 所示。

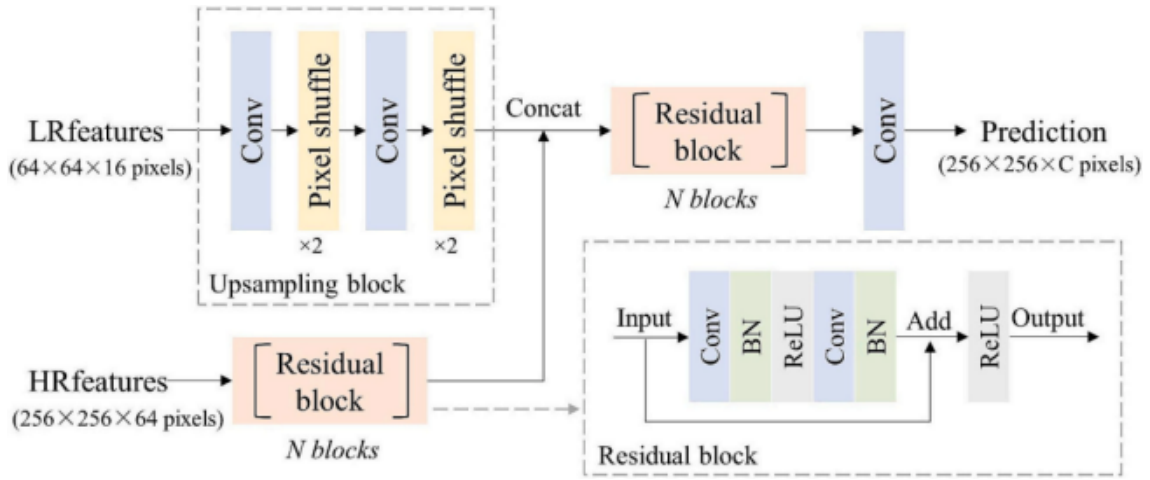


图 3. 高分辨率融合模块

3.4 不确定性感知损失函数

论文作者采用了不确定性感知损失函数来计算每个预测任务的权重。作者还在解码器的末尾添加了一个卷积层，用于估计低分辨率建筑物高度，以促进网络训练的收敛。每个任务的初始权重设置为 1.0，代表权重相等，在网络训练过程中会根据每个任务的预测不确定性逐步更新。不确定度加权损失的公式 2 如下所示：

$$L_{\text{uncertainty}} = \exp(-\log \sigma^2) \cdot L_{\text{original}} + \frac{\log \sigma^2}{2} \quad (2)$$

其中， L_{original} 为原始损失函数， σ 为一个可学习的参数。

4 复现细节

4.1 与已有开源代码对比

本文在 GitHub 上公开了源代码，本次的复现工作正是在该源代码的基础上进行修改和添加完成的，使用了论文中所提出的 Sentinel-1、Sentinel-2 和来自 Google Earth 的高分辨率卫星图像。

4.2 实验环境搭建

本次复现任务所使用的编程环境为 Python3.10, torch \geq 1.13.0, torchvision \geq 0.14.0, 基于 anaconda 的虚拟环境, 并且使用深度学习框架 Pytorch, 在 window 10 的 Pycharm 上进行代码编辑, 所有实验都是在 Tesla P100 上进行的。

5 实验结果分析

5.1 准确率估计

在该论文中选择了均方根误差 (RMSE)、平均绝对误差 (MAE) 和平均误差 (ME) 作为评估指标, 同时, 采用了总体准确率 (OA)、并集平均交并比 (mIoU) 以及频率加权交并比 (FWIoU) 来计算方法的精度, 如下表1展示了与作者论文中的准确率的比较。

表 1. 不同地区高度估计和高度分层的比较结果

地区	RMSE (m)	MAR (m)	ME (m)	OA (%)	MIoU (%)	FWIoU (%)
中国 ours	10.25	5.71	5.68	73.8	24.6	65.4
美国 ours	5.29	3.08	3.06	82.5	25.8	75.6
欧洲 ours	3.88	2.27	2.21	84.2	26.1	77.6
中国 author	10.318	6.086	3.926	73.8	24.6	65.4
美国 author	5.654	3.653	1.735	82.5	25.8	75.6
欧洲 author	4.113	2.589	1.539	84.2	26.1	77.6

对于高度估计所提出的方法在中国得到了最高的 RMSE (10.25)、在美国获得了第二高的 RMSE (5.29)、在欧洲获得了最低的 RMSE (3.88)。对于高度分层, 在中国获得了最低的 OA (73.8%)、在美国得到了第二高的 OA (82.5%)、在欧洲得到了最高的 OA (84.2%)。该方法在欧洲的表现最好、在美国次之、而在中国最差。在这三个地区中, 中国的高度分布异质性最为显著, 这让准确的高度预测更加具有挑战性, 而在美国和欧洲的建筑物空间分布更加集中, 这种分布更容易让模型捕捉。

该论文的准确率对比结果表明, 所提方法在不同地区的表现存在差异, 这表明未来可以通过针对性地调整算法参数或改进模型结构来优化性能, 以提高不同地区的准确率。

5.2 可视化结果

图 4显示了来自中国的测试样本上预测的建筑物高度和水平的示例。可以看出, 预测结果可以表现出从低层建筑到高层建筑的精细高度变化, 并且与参考结果具有高度一致性。此外, 可以看出预测结果包含比 Sentinel-1/2 图像更丰富的空间细节。

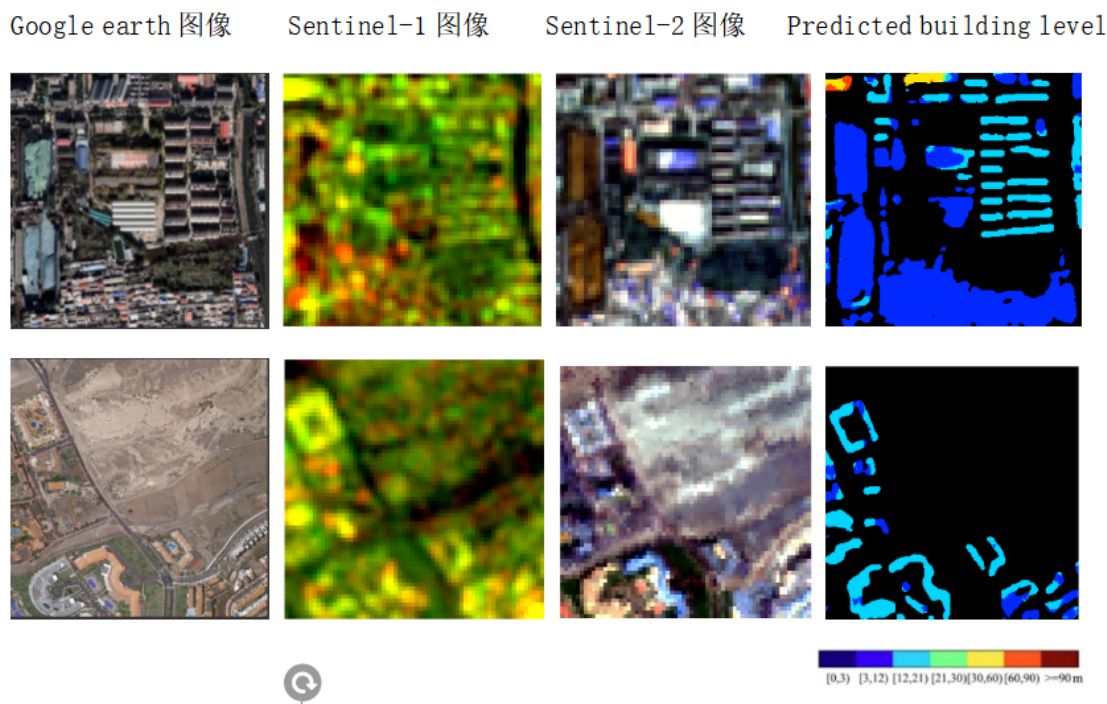


图 4. 可视化结果

6 总结与展望

通过复现一篇杰出的研究工作，我得以深入探索并投身于该研究领域。这一过程不仅是与前沿工作的一次深入交流，而且通过这次复现，我对领域内的相关研究和实践方法有了更加深刻的理解和认识。这次经历不仅锻炼了我的编程技能，更重要的是，它极大地提升了我的文献阅读和理解能力。

在复现过程中，我面临了诸多挑战，包括但不限于：其中代码的运行问题，这些成果是否采用了最先进的方法，能否理解每一步操作的深层含义，以及思考当前工作是否有进一步改进的空间、改进的方法和具体的实现途径。

尽管我已经成功实现代码，但遗憾的是，我还没有能够实现重要的框架创新。我相信在未来一定能够达到预期的研究目标，以及在此基础上进行创新，将是我后续工作的重点之一。通过解决代码运行中遇到的困难，我期待在后续的研究中能够实现真正的创新突破。

参考文献

- [1] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
- [2] Bowen Cai, Zhenfeng Shao, Xiao Huang, Xuechao Zhou, and Shenghui Fang. Deep learning-based building height mapping using sentinel-1 and sentinel-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 122:103399, 2023.

- [3] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.
- [4] Kelvin C.K. Chan, Xintao Wang, Xiangyu Xu, Jinwei Gu, and Chen Change Loy. Glean: Generative latent bank for large-factor image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14245–14254, June 2021.
- [5] Yunhao Chen, Jiatong Wu, Ke Yu, and Dandan Wang. Evaluating the impact of the building density and height on the block surface temperature. *Building and Environment*, 168:106493, 2020.
- [6] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [7] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 184–199, Cham, 2014. Springer International Publishing.
- [8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.
- [9] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [10] Huabing Huang, Peimin Chen, Xiaoqing Xu, Caixia Liu, Jie Wang, Chong Liu, Nicholas Clinton, and Peng Gong. Estimating building height in china from alos aw3d30. *ISPRS Journal of Photogrammetry and Remote Sensing*, 185:146–157, 2022.
- [11] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.
- [12] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

- [13] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1140, 2017.
- [14] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Unsupervised learning for real-world super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3408–3416, 2019.
- [15] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1874–1883, 2016.
- [16] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1905–1914, 2021.
- [17] Piper Wolters, Favyen Bastani, and Aniruddha Kembhavi. Zooming out on zooming in: Advancing super-resolution for remote sensing, 2023.
- [18] Wan-Ben Wu, Jun Ma, Ellen Banzhaf, Michael E. Meadows, Zhao-Wu Yu, Feng-Xiang Guo, Dhritiraj Sengupta, Xing-Xing Cai, and Bin Zhao. A first chinese building height estimate at 10 m resolution (cnbh-10 m) using multi-source earth observations and machine learning. *Remote Sensing of Environment*, 291:113578, 2023.