关于 DRCT 论文的复现

摘要

本文介绍了图像超分的相关算法，重点介绍了较为前沿的 DRCT 方法，复现了 DRCT 模型在法线图超分辨率任务上的应用。通过在 WPS+ 数据集上进行训练和 DiLiGenT 数据集上的测试，评估了模型的性能。实验结果表明，DRCT 模型在法线图超分辨率任务中表现出色，能够有效恢复高分辨率法线图。

关键词： 图像超分辨率；Transformer

1 引言

单图像超分辨率（SISR）旨在从给定的低分辨率（LR）图像中恢复高分辨率（HR）图像。相比于传统方法，卷积神经网络（CNN）取得了显著的成果 [3,4,10,15]，通过残差网络 [7] 和密集连接构 [9] 建了复杂的网络框架。然而 CNN 仍存在一些不足之处。首先，CNN 的局部感受野限制了其捕捉长距离依赖关系的能力，这在处理复杂图像结构时可能导致细节丢失。其次，CNN 的层次结构使得信息逐层传递，可能导致梯度消失或爆炸问题，影响训练效果，导致无法收敛。此外，CNN 在处理不同尺度的特征时表现不佳，难以同时捕捉全局和局部信息。

Transformer [17] 最初在自然语言处理(NLP)领域取得了显著成果，受此启发，Dosovitskiy 和 Alexey 提出了视觉 Transformer（ViT）[5]，将 Transformer 架构应用于计算机视觉任务。ViT 将输入图像分割成固定大小的非重叠块，并将它们作为序列进行处理。swinIR [12] 将 Swin Transformer [13] 应用于 SR 领域。

尽管基于 Transformer 的方法取得了显著的成果，但 Hsu 等人 [8] 发现随着网络深度的增加，特征图的强度分布会发生更大的变化，但在网络的末端通常会急剧下降，空间信息的丢失，表明存在信息瓶颈问题。因此，提出了 DRCT (Dense-residual-connected Transformer)，将信息向前传播以防止信息瓶颈，DRCT 能够以较少的参数和简化的模型架构实现更大的感受野。

1.1 文章的组织

本文的组织如下：在第二节介绍图像超分的相关工作，主要分 CNN、GAN [6] 和 Transformer 这三个方面介绍，在第三节介绍了本文的复现论文 DRCT，主要讲了论文的框架和训练策略和损失函数，在第四节中介绍复现细节，涉及数据库、度量指标、实验细节、与已有开源代码对比、创新点这五个点，在第五节展示了复现结果，在第六节中总结和展望。

2 相关工作

在本节中主要介绍了图像超分辨率领域的相关工作，具体分为基于卷积神经网络（CNN）、生成对抗网络（GAN）和 Transformer 的方法。

2.1 基于 CNN 的方法

卷积神经网络（CNNs）在图像超分辨率（SR）领域的发展中发挥了关键作用。在过去十年中，提出了许多基于 CNN 的方法。

SRCNN [3] 是将卷积神经网络应用于图像超分辨率问题领域的开创性工作之一。SRCNN 的网络架构由三个卷积层组成，随后是一个非线性激活函数。SRCNN 架构简单但相比于传统插值方法取得了显著的改进。然而这种简单的架构也限制了捕捉复杂图像细节和长距离依赖关系的能力。此外，SRCNN 执行像素级回归，有时可能导致模糊的结果，尤其是在大尺度上采样时。FSRCNN [4] 作为 SRCNN 的改进版本被提出，以解决后者在计算效率方面的不足。FSRCNN 直接将低分辨率图像映射到高分辨率图像，通过收缩扩张方法来减少网络的参数，降低计算复杂度。ESPCN [15] 提出了亚像素卷积层，允许网络直接从低分辨率输入生成最终的高分辨率图像，无需进行单独的上采样操作，这种方法通过显著减少上采样过程中涉及的参数和操作数量，提高了计算效率。随着残差网络 [7] 的成功，Kim 等人 [10] 将其应用于 SR 领域提出了 VDSR，其具有 20 卷积层，具有非常大的非常大的感受野，通过学习 LR 和 HR 图像之间的残差从而更快地收敛。

2.2 基于 GAN 的方法

生成对抗网络（GANs）[6] 在图像超分辨率（SR）中的应用显著推动了该领域的发展，解决了与基于传统 CNN 方法相关联的许多局限性。GANs 利用对抗训练机制，其中生成器网络学习生成高分辨率图像，判别器网络区分真实和生成的高分辨率图像。这一对抗过程鼓励生成器产生比先前方法更真实纹理和更清晰细节的图像。

SRGAN [11] 是首次将 GAN 应用于图像超分辨率任务的工作之一。其在目标函数中使用感知损失。这种损失函数从预训练网络（例如 VGG 网络 [16]）中提取的高级特征。Wang 等人 [19] 在 SRGAN 的基础上提出了 ESRGAN，使用残差-残差密集块（RRDB）学习更复杂的表示，提出了相对平均判别器（RAD），更多地关注细粒度细节，从而进一步提高生成图像的感知质量。为解决现实世界的图像超分辨率问题，如伪影、噪声和低质量，Wang 等人 [18] 在 ESRGAN 的提出了 Real-ESRGAN，采用真实世界的退化图像作为训练数据，而非通过下采样获得的合成低分辨率图像，从而捕捉各种形式的失真。

2.3 基于 Transformer 的方法

Transformer 最初是为自然语言处理任务设计的，最近已成为计算机视觉领域的强大工具。相比于 CNN，Transformer 捕捉长距离依赖关系和建模数据内部复杂交互的能力。

视觉 Transformer（ViT）是将 Transformer 架构应用于计算机视觉任务的开创性工作之一。ViT 将输入图像分割成固定大小的非重叠块，并将它们作为序列进行处理，Transformer

通过自注意力机制关注这些标记来学习空间关系。然而 ViT 将图像分成固定的块，这会导致块边界产生伪影，每个块独立处理，没有信息交流。

Liu 等人 [13] 提出了 Swin Transformer, 将注意力限制在非重叠的局部窗口，并在层之间进行位移，使注意力机制更高效，同时保留全局上下文。Liang 等人 [12] 在 Swin Transformer 应用于 SR 领域，提出了 SwinIR，通过残差 Swin Transformer 块 (RSTB) 进行深度特征提取。HAT [1] 将通道注意力机制与重叠交叉注意力模块相结合，更好地聚合跨窗口信息。

3 DRCT

3.1 网络架构

DRCT 由浅层特征提取、深层特征提取和图像重建三个模块组成，如图1所示

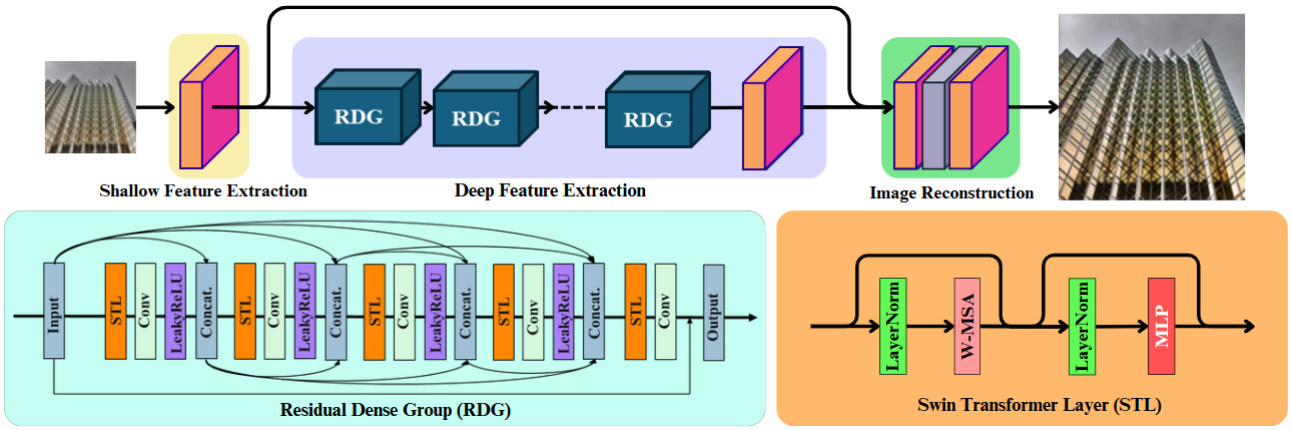

图 1. 整体框架图来源 DRCT 论文 [8] 中的图片。

浅层特征主要包含低频部分，深层特征则专注于恢复丢失的高频部分，图像重建将浅层特征和深层特征转换为图像。

3.1.1 浅层特征提取

浅层特征提取由 3×3 的卷积层 $Conv(\cdot)$ 构成，从低分辨率图 $I_{LQ} \in \mathbb{R}^{H \times W \times C_{in}}$（$H$、$W$ 和 C_{in} 分别是图像高度、宽度和输入通道数）提取浅层特征 $F_0 \in \mathbb{R}^{H \times W \times C}$，即

$$\mathbf{F}_0 = \mathrm{Conv}\left(\mathbf{I}_{LQ}\right), \tag{1}$$

相比于 Transformer，卷积层擅长早期视觉处理，导致更稳定的优化和更好的结果，并将输入图像空间映射到更高维的特征空间。

3.1.2 深层特征提取

深层特征则专注于恢复丢失的高频部分，由 K 个残差密集组（Residual-Dense Group，RDG）组成和一个卷积层组成，即

$$\begin{aligned} F_i &= \mathrm{RDG}_i(F_{i-1}), \quad i = 1, 2, \ldots, K, \\ F_{DF} &= \mathrm{Conv}(F_K), \end{aligned} \tag{2}$$

其中，RDG_i 表示第 i 个残差密集组，Conv 是最后一个卷积层，添加卷积层有助于 Transformer 关注局部结构特征和依赖关系，提高其从输入中捕获相关信息的能力，使得模型更好地融合浅层和深层地不同层次的特征。

基于残差密集网络 RDN [20]，残差密集组 RDG 集成 Swin-Transformer 层（STL）和密集残差连接，充分融合、利用多层次特征，由密集连接层 + 局部特征融合（LFF）+ 局部残差构成，即

$$\begin{aligned} Z_j &= H_{\text{trans}}\left(\mathrm{STL}\left([Z, \ldots, Z_{j-1}]\right)\right), \quad j = 1, 2, 3, 4, 5, \\ \mathrm{SDRCB}(Z) &= \alpha \cdot Z_5 + Z, \end{aligned} \tag{3}$$

对于 RDG 的每个中间特征 Z_j，通过密集连接 $[\cdot]$ 输入到后续的每一层中，每一层通过 concat 整合前面层的所有特征，通过 1×1 的卷积将这些不同层次的特征自适应融合，$H_{trans}(\cdot)$ 表示具有 LeakyReLU 激活函数的卷积层，通过残差，将第 $i-1$ 个 RDG 的输出 Z 和第 i 个 RDG 的融合特征 Z_5 相加，其中 α 表示残差缩放因子。

3.1.3 图像重建

使用全局残差连接来融合浅层特征和深层特征，再采用亚像素卷积 PixelShuffle [15] 将特征图从低分辨空间转换到高分辨空间中，即

$$\mathbf{I}_{SR} = H_{\text{rec}}(\mathbf{F}_0 + \mathbf{F}_{DF}) \tag{4}$$

其中 $H_{rec}(\cdot)$ 表示重建函数，在文章中采用 PixelShuffle 方法，将输入特征 $F \in \mathbb{R}^{C \times h \times w}$ 进行特征提取，生成含 $r \times r$ 个通道的特征图，最后生成上采样结果 $I_{SR} \in \mathbb{R}^{C \times hr \times wr}$，

3.2 训练策略和损失函数

DRCT 采用同一任务渐进式训练策略（Same-task Progressive Training Strategy，PTS），首先在 ImageNet [2] 上预训练 DRCT 以初始化模型参数，在特定数据集上使用 L1 损失，

$$\ell_{L1} = \|I_{HR} - I_{SR}\|_1 \tag{5}$$

最后使用 L2 损失，损失消除孤立像素和伪影，

$$\ell_{L2} = \|I_{HR} - I_{SR}\|_2 \tag{6}$$

4 复现细节

4.1 数据集

DRCT 模型在 WPS+（wonderful photometric stereo plus）数据集上训练，WPS+ 是一个基于光度立体方法的法线图数据集，其包含 600 个对象，测试阶段采用 DiLiGenT 数据集 [14] 和 WPS+ 数据集，通过应用双三次降采样方法生成 x4 的低分辨率图像。

4.2 度量指标

为了进行定量比较，使用 SISR 任务中常用的指标：峰值信噪比（PSNR）和结构相似性指数（SSIM），

$$\text{PSNR} = 10 \times \log_{10}\left(\frac{\text{MAX}_I^2}{\text{MSE}}\right) \tag{7}$$

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{8}$$

此外，使用平均角误差（MAE），

$$MAE = \frac{1}{|\mathbf{N}|}\sum_{i,j} arccos\left(\tilde{\mathbf{n}}_{i,j} \cdot \mathbf{n}_{i,j}\right), \tag{9}$$

来定量测量 3D 重建结果，其中 $\tilde{\mathbf{n}}_{i,j}, \mathbf{n}_{i,j}$ 分别表示预测的法线和真实法线，$|\mathbf{N}|$ 表示输入法线像素的总数，MAE 计算角误差的均值 MEAN、中位数 MID、方差 VAR、5 度角误差内的像素百分比、10 度角误差内的像素百分比这五个指标，前三者越高表示重建质量越好，后两者越低表示重建质量越好。

4.3 实验细节

训练采用使用 Adam 优化器，$_1 = 0.9, _2 = 0.99$，批大小设置为 32，块大小为 128，应用权重衰减 0.999，过程分为两个阶段，第一阶段使用 L1 损失训练训练 50k 次迭代，学习率初始化为 2×10^{-4} 使用多步学习调度器, 学习率分别在 25k、40k、45k、47.5k 次迭代时减半，第二阶段采用 L2 损失训练 25k 次迭代，学习率初始化为 1×10^{-5} 使用多步学习调度器, 学习率分别在 12.5k, 20k, 22.5k, 24k 次迭代时减半。

本实验还测试了在作者提供的 ImageNet 预训练模型（采用 L1 损失函数训练了 800k 迭代次数）上进行微调，采用使用 L1 损失训练 25k 次迭代，

为了提高泛化能力，在图像读取时应用随机水平翻转和旋转增强。

4.4 与已有开源代码对比

本复现实验参考的代码在https://github.com/ming053l/DRCT获取，同时该代码是基于basicSR框架。

在 DRCT 的网络架构中，我修改了图像归一化的相关代码，原模型是在 DIV2K 图像的超分，因此其减去 DIV2K 数据集的均值 (0.4488, 0.4371, 0.4040)，而对于法线图其在非 mask 区域内的均值应为 (0.5，0.5，0.75)（法线的 z 轴是面向相机的，其值大于 0.5），然而对于不同法线图，由于其形状各异，其 mask 区域占比不一，具有较大的差异，因此在本实验中对于了采用了 (0,0,0) 和 (0.5，0.5，0.75) 这两种不同均值处理的方法。

在度量的相关代码中，添加了关于法线图的度量指标，计算了角误差的均值 MEAN、中位数 MID、方差 VAR、5 度角误差内的像素百分比、10 度角误差内的像素百分比这五个指标。

在训练管道上修改了 basicSR 框架的代码，原本 basicSR 框架对于验证集和测试集是每次读取一张图片即批大小为 1，numworker 也为 1，我通过修改 basicSR 关于创建数据加载器

的相关代码，使得对于验证集和测试集能够能够自定义批大小和 numworker 从而加快图像的读取和恢复。

重写了一个用于法线图读取的 dataset，与彩色图相比，法线图在读取时不需要进行从 bgr 到 rgb 的转换，其本身存储的通道顺序即为 rgb 对应这 xyz 这三个方向的法向，在保存图片时也不需要从 rgb 转换到 bgr，直接以 rgb 通道存储。

为了加快数据的读取，将 2k 的图像裁剪为重叠区域为 128，256*256 的大小的子图，同时将图像从 png 格式转化为 lmdb 格式的二进制特殊格式以加快法线图的读取。

4.5 创新点

将传统的 RGB 数据集替换为法线图数据集，测试该模型在法线图超分上的性能。

法线图和传统的 RGB 图像在数据分布上存在不同，RGB 图像在均值为 (0.5,0.5,0.5) 附近，而法线图的非 mask 区域的均值在 (0.5，0.5，0.75) 附近，复现实验探讨了在 ImageNet 上预训练，并在 WPS+ 数据集上微调的可能性。

5 实验结果

本实验测试了图像减去 (0.5,0.5,0.75) 和 (0,0,0)（即不处理）这两种数据处理方法，比较了两种训练方式，一种是使用在 ImageNet 上的预训练模型，一种是使用 WPS 数据集从头开始训练，对 WPS 数据集和 DiLIGenT 数据集上进行测试。最终的结果如表1所示。

结合表1分析可知，采用 ImageNet 预训练能够在 DiLiGenT 数据上取得最好的结果，这是由于模型能够在常规彩色图上学习到通用的特征和分布，且 ImageNet 数据集相比于 WPS 数据集含有更多的数据，能够从中学到更多的先验知识。相比于从头开始训练的 DiLiGent 数据集上的测试，其他情况下采用数据预处理都会得到更好的结果，然而对于使用 DiLiGenT 数据集，其使用数据预处理和不使用数据预处理，在 PSNR 上相差 2 个点，其效果甚至不如仅使用预训练模型。

L2 损失函数能够很好地消除伪影，如图2对比了 DiLiGenT 上 ball 法线图的修复情况，可以看出相比于使用 L1 损失函数使用 L2 损失函数得到的边界更光滑、流畅，相比于使用归一化，不使用归一化会在窗口边界形成伪影。

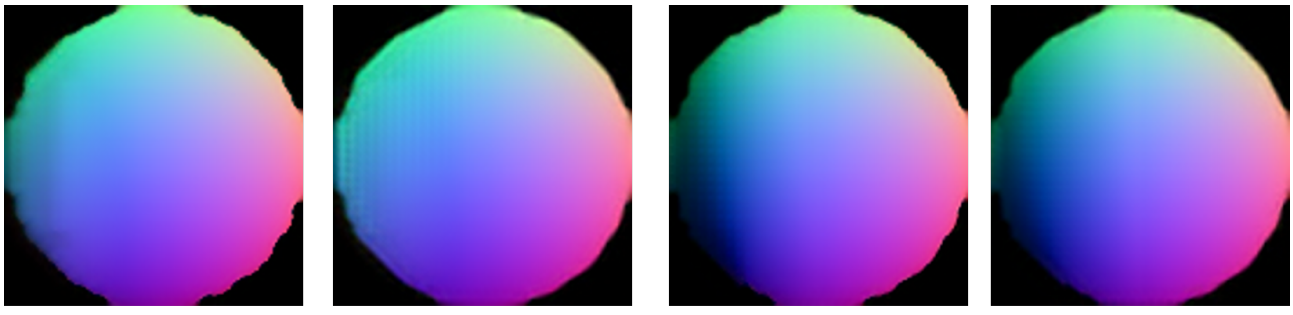

图 2. DiLiGenT 数据集中的 ball 在不同方法下的对比。(a) 为不使用归一化，采用 L1 损失函数训练的结果，(b) 为不使用归一化，采用 L2 损失函数继续训练的结果，(c) 为使用归一化，采用 L1 损失函数训练的结果，(d) 为使用归一化，采用 L2 损失函数继续训练的结果。

表 1. 不同方法在 WPS 和 DiLIGenT 数据集上的对比

数据集	训练方式	数据处理	阶段	PSNR[+]	SSIM[+]	MEAN[-]	MID[-]	VAR[-]	5°[+]	10°[+]
DiLi GenT	从头开始	否	1	22.6804	0.8574	7.6154	3.441	165.28	0.607	0.7619
			2	23.2961	0.8565	7.8814	3.5502	149.86	0.598	0.7362
		是	1	20.0716	0.8292	12.7587	4.737	313.32	0.525	0.646
			2	20.5079	0.8322	12.8703	5.558	271.45	0.49	0.619
	预训练 + 微调	否	1	20.9922	0.8123	7.9207	**2.995**	215.39	0.672	0.7841
			2	24.4423	0.8811	6.1478	3.349	121.37	0.723	0.8708
		是	1	20.1951	0.7929	6.1657	3.145	**94.784**	0.707	0.835
			2	**24.5913**	**0.8868**	**5.9046**	3.105	117.86	**0.749**	**0.8744**
WPS	从头开始	否	1	31.5035	0.8958	3.086	2.3822	21.1808	0.829	0.9592
			2	32.2498	0.8959	3.1134	2.3931	19.5982	0.827	0.9581
		是	1	31.8626	0.8973	**3.0578**	**2.37**	19.077	0.83	**0.9596**
			2	32.3642	0.8972	3.0865	2.381	19.001	**0.829**	0.9588
	预训练 + 微调	否	1	29.107	0.863	4.1185	3.064	30.904	0.747	0.9156
			2	32.4447	0.8958	3.0938	2.4	18.202	0.828	0.9582
		是	1	28.0736	0.8518	3.6477	3.008	**10.863**	0.782	0.9437
			2	**32.4931**	**0.8963**	3.085	2.396	18.007	0.828	0.9585

表中显示了各种方法在 DiLiGenT 和 WPS 数据集上的关于 PSNR、SSIM、MEAN、MID、VAR、5°、10°7 个度量指标，其中加号'+' 表示越大越恢复效果越好，减号'-' 表示越小恢复效果越好。对于了从头开始在 WPS 数据集上训练和使用作者提供的在 IgameNet 上预训练其后在 WPS 数据集上微调的两种训练方式。数据预处理有两种，“否” 表示不适用数据归一化，“是” 表示使用数据归一化。对于从头开始训练，阶段 1 表示使用 L1 损失函数训练，阶段 2 表示在阶段 1 的基础上使用 L2 损失函数继续训练。对于预训练和微调，阶段 1 表示直接使用在 ImageNet 上训练的模型，阶段 2 表示在预训练模型基础上使用 L2 损失函数训练。其中黑色粗体表示当前度量指标下的最好的方法。

6 总结与展望

6.1 总结

本研究通过复现实验，深入测试了 DRCT 模型在法线图超分辨率任务上的应用效果。实验结果基本满足了预定的实验要求，证实了模型在特定条件下的有效性。然而，本复现实验仍存在若干不足之处：

1. **缺乏对比分析：** 本研究未能将 DRCT 模型与其他方法进行直接的性能对比。这限制了对 DRCT 模型性能的全面评估。

2. **缩放因子的局限性：** 实验仅在 $\times 4$ 缩放尺度上进行了测试。未能全面展示模型在不同缩放条件下的泛化能力，可能限制了对模型鲁棒性的评价。

3. **训练次数与数据集规模的较小：** 本在训练次数方面，复现实验在法线图上的训练次数相

较于在 ImageNet 上的训练次数减少了近一个数量级，在训练数据集规模上，ImageNet 数据集包含约 1400 万张图片，而 WPS+ 数据集仅包含 500 张图片。

6.2 展望

图像超分辨率领域已取得诸多突破性进展，这些成果为法线图超分辨率任务提供了宝贵的参考与启示。鉴于法线图的独特性质，本实验采用的 L1 损失和 L2 损失可能并非最优选择。未来的工作可考虑开发或调整更适合法线图特性的损失函数，以进一步提升模型性能。

参考文献

[1] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22367–22377, 2023.

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[3] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.

[4] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 391–407. Springer, 2016.

[5] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[8] Chih-Chung Hsu, Chia-Ming Lee, and Yi-Shiuan Chou. Drct: Saving image super-resolution away from information bottleneck. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 6133–6142, June 2024.

[9] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[10] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

[11] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[12] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021.

[13] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[14] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3707–3716, 2016.

[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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

[16] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[17] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

[18] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021.

[19] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In

Proceedings of the European conference on computer vision (ECCV) workshops, pages 0–0, 2018.

[20] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018.