高级语言程序设计

实验报告

学院 计算机学院

班级 计算机科学卓越班

学号 2313903

姓名 付嘉晨

2024 年 5 月 13 日

目录

1 作业题目

1.1 中文题目

集成了用于图像重建的多尺度感知神经网络的图片编辑器

1.2 英文题目

Rethinking Pixel-level Predictions with Multi-scale Features: An Image Editor Integrating Multi-scale Perceptual Neural Networks for Image Reconstruction

2 开发环境

Visual Studio2022、Qt Creator 5.0.2、OpenCV4.5.5、LibTorch1.12.0

3 背景介绍

3.1 神经网络

以深度学习模型为代表的神经网络已经成为计算机领域的一项革命性技术，其主要思想为使用神经网络层如**密集连接层 (Dense or Linear)**、**卷积层 (Convolution)** 等及激活函数构造复杂的模型，通过对数据集进行拟合来进行表征的学习。在计算机视觉领域中，常使用的神经网络模型有以下几种：

卷积神经网络 LeCun 等人 [1] 首次将卷积模型用于手写数字识别，而 Krizhevsky 等人 [2] 使用 ImageNet 数据集首次训练出深度卷积网络，并在 ImageNet 竞赛中取得了突破性的胜利。在之后，牛津大学的视觉几何实验室提出了经典分块网络 VGG[3]，深刻影响后来的网络设计；Kaiming 等人 [4] 于 2016 年提出残差网络，利用残差跳跃连接首次训练出上千层深度的网络，这样的残差结构在后来的众多深度学习模型中均被使用。在 2020 年之后，卷积网络的研究热度有所降低，但也有许多有意义的研究出现，这其中，令人影响深刻的有清华大学丁霄汉等人提出的 RepVGG[5]，使用结构重参数化技术重构了 VGG[3] 网络；同样由清华大学丁霄汉等人提出的 RepLKNet[6]，发掘了大核卷积的潜力；FAIR 提出的 ConvNeXt[7]，总结了多年以来的训练经验以及技巧，训练出了性能极佳的卷积神经网络。

Transformer 网络 人类在看到一张图片或者阅读一段文字时，并不是从头到尾顺序地获取信息，而是基于注意力地进行感知。长期以来，学者们都尝试让网络获取注意力，这称之为注意力机制，但学界对于注意力机制的使用长期停留在对网络 “锦上添花”，即在主干网络之后添加注意力机制以获得增强的特征信息。2017 年，谷歌的研究团队创新性

地提出了完全基于注意力机制的用于机器翻译任务的神经网络模型 Transformer[8]，这样的架构设计点起了自然语言处理学界的研究热潮，而计算机视觉学界的研究者们也不断在思考如何将这样一种结构引入视觉任务中，但由于图像的二次复杂度以及低级语义特征，使得像自然语言处理那样的上下文处理变得困难。2020 年，同样来自谷歌的团队提出了 Vision Transformer[9]，首次将 Transformer 架构引入视觉任务；在之后，微软亚洲研究院的团队对 ViT 做出了改进，引入了滑动窗口机制，提出 Swin Transformer[10]，进一步提高了神经网络在 ImageNet 图像分类任务上的准确度，并且为后来的研究提供了思路。

3.2 图像超分辨率

图像超分辨率是一个重要的底层视觉（Low-Level Vision）任务，其目的是从低分辨率的图像中重建出高质量的版本。根据输入的不同可以分为两个子任务：**单一图像超分辨率（SISR）**与**基于多幅图像的超分辨率（MISR）**，本工作聚焦于 SISR 这一任务，并使用深度学习的方法实现超分辨率。多年来，基于深度学习的超分辨率方法主要有以下几类：

基于卷积神经网络的图像超分辨率 自 Dong 等人 [11] 首次将神经网络引入图像超分辨率任务并取得较好的成果以来，大量研究者聚焦于如何使用神经网络进一步提升超分辨率的性能，这其中，有里程碑意义的有：2016 年由 Shi 等人 [12] 提出的 ESPCN 网络，他们引入了**亚像素卷积**这一操作，使图像超分辨率的计算复杂度大幅下降；2017 年由 Lim 等人 [13] 提出的增强深度超分辨率网络，删除了传统网络中非必要的模块以及冗余的操作，在使模型轻量化的同时提升了超分辨率的效果；2018 年由 Zhang 等人 [14] 提出的残差通道注意力网络，首次将注意力机制引入图像超分辨率任务中，在效果比 EDSR[13] 好的同时仅使用了其 $\frac{1}{3}$ 的参数量，证明了注意力机制对超分辨率任务的有效性。

基于生成对抗网络的图像超分辨率 生成对抗网络于 2014 年由 Ian 等人 [15] 提出，其生成图片的能力令学界震惊。2017 年，Ledig 等人 [16] 首次将 GAN[15] 引入图像超分辨率任务中，并且提出了**内容损失**，使整个网络在学习的过程中更加关注重建图像和原始图像的语义特征差异，而非逐个像素之间的颜色和亮度差异。在之后，Wang 等人 [17] 提出的 ESRGAN 网络进一步提高了性能，取得了不错的成绩。

基于 Transformer 的图像超分辨率 2020 年，Yang 等人 [18] 首次将 Transformer[8] 引入图像超分辨率任务中。在之后，Liang 等人 [19] 于 2021 年提出 SwinIR 网络，将 Swin Transformer[10] 引入超分辨率任务，刷新了图像超分辨率的 SOTA（State-Of-The-Art）。2023 年，南开大学 Zhou 等人 [20] 提出 SwinIR[19] 的改进网络 SRFormer，进一步提升了网络的性能。

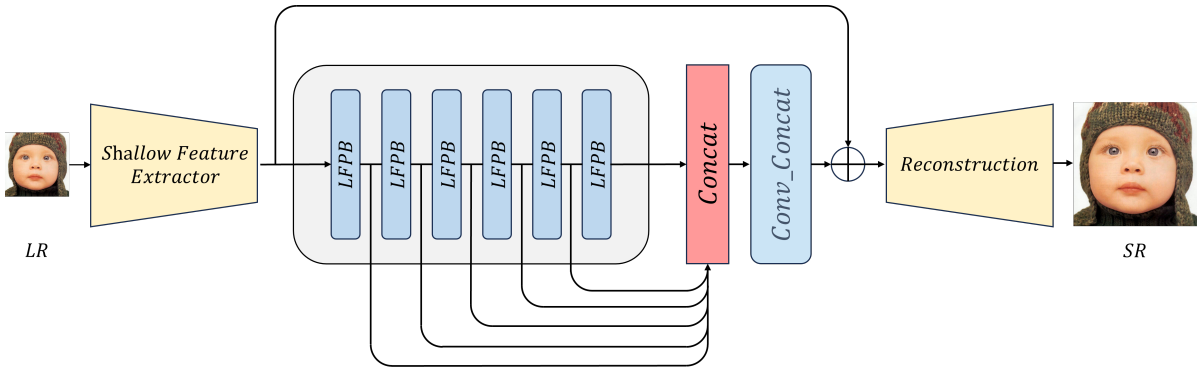

图 1: 多尺度感知神经网络结构图.

基于轻量级神经网络的图像超分辨率 神经网络的特征提取与表达能力令人震惊的同时，如何将其有效地部署备受学界及工业界的思考，这其中，如何将网络用在计算能力较差的移动端设备上是极具现实意义的问题。由于图像处理的二次复杂度以及深度卷积网络的大参数量，使得像 RCAN[14] 这样参数量达到 10M 级别的网络难以用于移动端，而 Transformer 网络如 SwinIR[19] 由于使用了大量的多头自注意力，产生了极大的并行负担，也难以被移动端直接使用。2017 年，Howard 等人 [21] 将深度可分离卷积用于神经网络结构中，大幅降低网络参数量的同时，性能衰减几乎可以忽略不计，给学界提供了一种轻量化模型的思路；2020 年，Haase 等人 [22] 基于对 Howard 等人工作的思考与实验，提出了一种深度可分离卷积的改进版本，蓝图卷积，在参数量小幅度增加的同时，有效地提升了深度可分离卷积的性能。2018 年，Zheng 等人 [23] 提出了信息蒸馏网络，使用一种由多分支架构组成的主干网络来提供蒸馏的语义信息，为轻量级超分辨率的网络设计提供了一种模板级思路；这之后，同样是 Zheng 等人 [24] 基于这一工作提出了改进版本 IMDN，进一步为网络引入多级蒸馏信息，提升了网络性能。2022 年，Li 等人 [25] 基于蓝图卷积 [22] 提出了用于高效率图片超分辨率的卷积神经网络，在参数量仅有 0.3M 的基础下取得了极其优秀的成绩；Kong 等人 [26] 提出 RLFN 网络，去除了先前广为使用的多分支结构，使网络推理提速。

4 研究动机

4.1 多尺度特征对于视觉任务的意义

南开大学的程明明教授在《粒度自适应的图像感知技术》[27] 中提出了这样的一个例子：如果只看一张大图的一小部分，不论是人或是机器都难以准确辨别出物体的类别，而在获得这个小部分周围的信息后，我们能轻易地辨别出这个物体是一个树桩样子的椅子。这充分说明了多粒度感知与上下文信息对于图像感知的重要性。同时，程教授提出了对 ResNet[4] 有效性的另一种思考，即：残差跳跃连接能够使小感受野的特征图与大感受野的特征图混合，给网络提供一种多粒度的信息，有利于图像感知。

而在图像分类、目标检测等高级视觉任务中，由于输入和输出可能分辨率不同、维

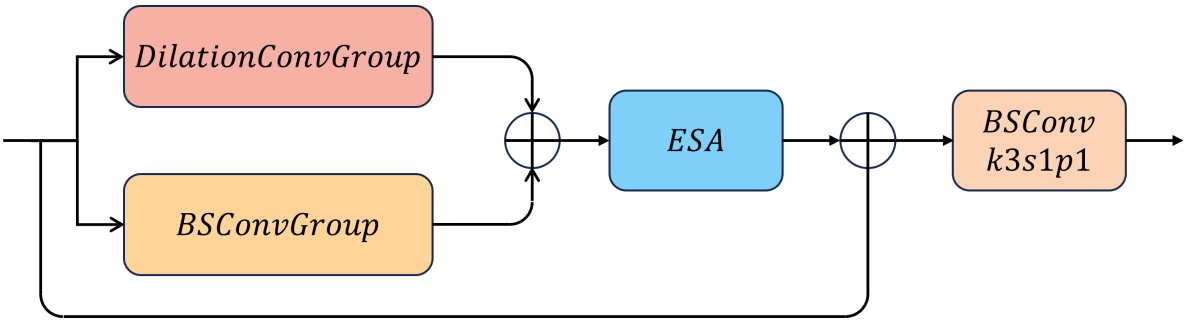

图 2: 多尺度感知神经网络块 LFPB.

度不同，因此可以引入池化等下采样操作来降低特征图的分辨率，一方面可以减小运算复杂度，另一方面可以对特征图中的信息进行融合与筛选，分辨率不同的特征图组成了“多粒度”的信息。而在超分辨率任务中，由于网络需要做的是上采样工作，因此在网络中的降采样操作可能是对模型性能有害的，因此分辨率不同的金字塔样“多粒度”信息提取对于这一任务可能不适用。但我们可以设计一种方法在保持特征图分辨率不变的条件下尽可能地提供不同尺度的信息，以达到“多粒度”的感知效果。出于“提供多尺度感知信息”的目的，我设计了这样一种结构与 IMDN[24] 相似的网络，并使用感受野不同的卷积组合来进行特征的提取，并提供多次特征图融合以混合多尺度信息。这样的网络设计具备从小到大几乎所有尺度的感知信息，实验证明，这样的网络效果十分不错。

4.2 如何在低参数量的条件下获得大的感受野

Chen 等人 [28] 提出的 Deeplab 架构在另一个像素级预测任务：**语义分割**上取得了极为优秀的成果，他们引入了空洞卷积这一卷积层设计，在保持卷积网络参数量不变的同时，提供了更大的感受野，这启发了我们的工作。但同时也有人认为，空洞卷积这一有像素跨越的操作对于像素密集型预测有不良的影响，而超分辨率正是一个像素密集预测的任务。

出于对空洞卷积是否对超分辨率有不良影响的思考，以及为网络提供大感受野的需求，我在网络中引入了空洞卷积。Wang 等人 [29] 的工作说明了如果整个网络都使用有空洞的卷积，那么一定会导致相邻像素的信息无法构建联系，因此，我们并非全部使用带有膨胀的卷积，而是使用膨胀系数 $dilation = [1, 2, 2]$ 的卷积组合，这一组合能提供 11×11 大小的感受野。同时，我们的网络块还包含一个由 3 个卷积核大小为 3×3 的蓝图卷积 [22] 组成的卷积组合，这一组合能提供 7×7 大小的感受野。

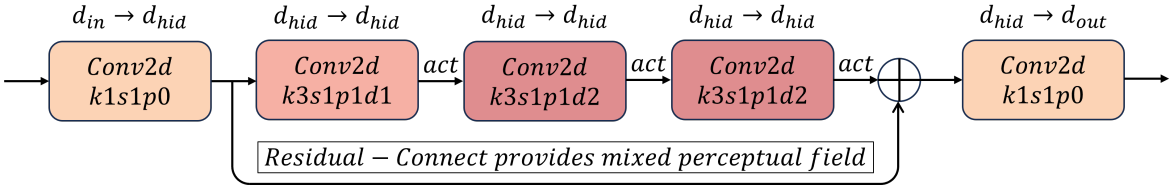

图 3: 膨胀卷积组合.

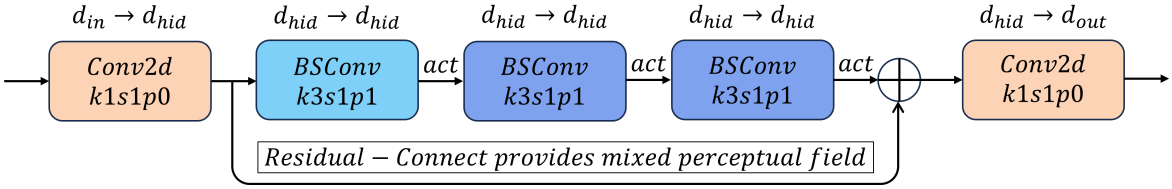

图 4: 蓝图卷积组合.

5 主要流程

5.1 模型训练

我们将上述模型在 DIV2K 数据集上进行训练, 训练一共进行了 1×10^6 个 iter, 训练的批次大小 $batch = 16$, 训练使用的优化器是 Adam 优化器, 初始学习率为 2×10^{-4}, 并使用余弦退火并周期性重启的方式进行训练, 重启周期为 $range = [3e5, 2e5, 2e5, 2e5, 1e5]$, 重启权重为 $weight = [1, 0.5, 1, 0.5, 0.5]$, 训练的真实图像 $PatchSize = 96$。

5.2 C++ 部署

我们将模型部署于 C++ 环境并进行测试。

参考文献

[1] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

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

[4] 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.

[5] Xiaohan Ding, Xiangyu Zhang, Ningning Ma, Jungong Han, Guiguang Ding, and Jian Sun. Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13733–13742, 2021.

[6] Xiaohan Ding, Xiangyu Zhang, Jungong Han, and Guiguang Ding. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11963–11975, 2022.

[7] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986, 2022.

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[10] 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.

[11] 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.

[12] 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.

[13] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

[14] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018.

[15] 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.

[16] 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.

[17] 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.

[18] Fuzhi Yang, Huan Yang, Jianlong Fu, Hongtao Lu, and Baining Guo. Learning texture transformer network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5791–5800, 2020.

[19] 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.

[20] Yupeng Zhou, Zhen Li, Chun-Le Guo, Song Bai, Ming-Ming Cheng, and Qibin Hou. Srformer: Permuted self-attention for single image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12780–12791, 2023.

[21] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[22] Daniel Haase and Manuel Amthor. Rethinking depthwise separable convolutions: How intra-kernel correlations lead to improved mobilenets. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14600–14609, 2020.

[23] Zheng Hui, Xiumei Wang, and Xinbo Gao. Fast and accurate single image super-resolution via information distillation network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 723–731, 2018.

[24] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th acm international conference on multimedia*, pages 2024–2032, 2019.

[25] Zheyuan Li, Yingqi Liu, Xiangyu Chen, Haoming Cai, Jinjin Gu, Yu Qiao, and Chao Dong. Blueprint separable residual network for efficient image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 833–843, 2022.

[26] Fangyuan Kong, Mingxi Li, Songwei Liu, Ding Liu, Jingwen He, Yang Bai, Fangmin Chen, and Lean Fu. Residual local feature network for efficient super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 766–776, 2022.

[27] Mingming Cheng. 粒度自适应的图像感知技术. https://mmcheng.net/report/.

[28] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[29] Panqu Wang, Pengfei Chen, Ye Yuan, Ding Liu, Zehua Huang, Xiaodi Hou, and Garrison Cottrell. Understanding convolution for semantic segmentation. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 1451–1460. Ieee, 2018.