B1 基于 MSR-net 的图像增强网络

本案例以 Shen[1] 等人的论文 MSR-net:Low-light Image Enhancement Using Deep Convolutional Network 为基础，介绍了基于深度学习的 MSR-net 图像增强的基本原理和实现流程，对该论文进行算法复现，并对代码进行了解析，指导学生完成图像增强实验、分析实验结果。通过本案例的学习，可以拓展视野，了解图像增强和深度学习领域前沿成果；通过阅读和复现优秀学术论文，培养学生自主学习的能力、工程实践能力和分析总结能力，提升科研兴趣和创新意识。

1 学习目标

⑴ 理解深度学习在图像增强领域中的应用；

⑵ 阅读文献、调试运行开源程序、分析对比实验结果，培养自主学习、研究的能力。

2 案例背景

图像增强是一种提高图像质量和信息量的技术。目的是针对图像应用的不同场合，增强图像中的有用信息，改善图像视觉效果，具有十分重要的应用价值和科研意义。

由于图像在采集过程中会受到天气、光照强度、光源方向等因素影响，导致所采集图像质量不高。近年来，大量的基于 Retinex 理论的传统图像增强算法被提出，应用于图像的增强处理，例如 SSR、MSR、MSRCR 等。随着 CNN 的出现，众多学者尝试将卷积神经网络与图像增强相结合，希望借助它强大的自主学习能力来解决传统方法无法解决的图像增强问题。

2017 年 Shen 等人发表的 MSR-net:Low-light Image Enhancement Using Deep Convolutional Network，提出传统的 MSR 算法模型和 CNN 模型相似，并首次将传统的 MSR 和 CNN 相结合，提出一种端到端的 MSR-net 网络，为后续卷积神经网络在图像增强领域的发展提供了基础。经过 MSR-net 增强后的图像亮度提升明显，同时也解决了色彩不自然问题。本案例以此论文为基础，介绍基于深度学习的 MSR-Net 图像增强算法的原理、方法以及算法的复现，并分析、对比实验结果。

3 MSR-net 原理

MSR-net 流程共分为三个部分：多尺度对数变换、卷积差分和颜色恢复。分别使用三个子函数来对应这三个部分，则 MSR-net 模型可以用函数表示为 $Y = \mathrm{f}(x) = f_3(f_2(f_1(X)))$，其中 X 为原始输入的微光图像，Y 为模型输出的增强后的图像，其中 f_{1-3} 分别表示网络中的三个过程。其网络结构如图 B1.1 所示。

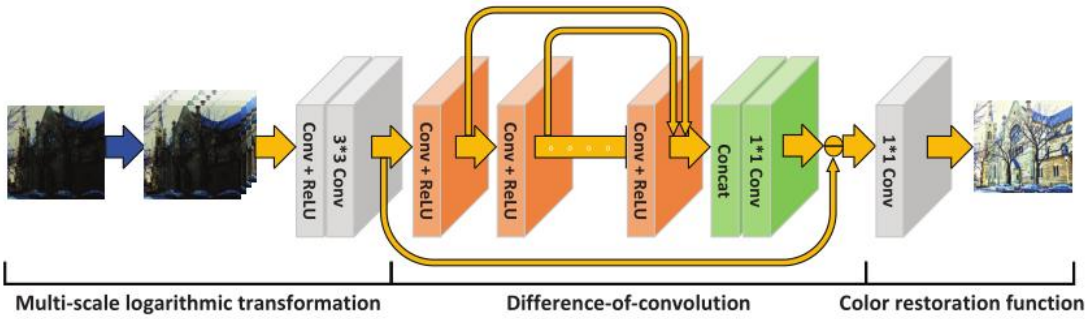


图 B1.1 MSR-net 网络结构图

3.1 多尺度对数变换

多尺度对数变换以原始微光图像 X 作为输入，计算出相同大小的输出 X1。首先，对暗图像进行 n 次差分对数变换增强，目的是增强暗图像，得到 n 个不同的图像。将这 n 个图像在通道维度上进行堆叠，得到一个 3n×宽×高的三维张量，然后经过 1 个 conv+ReLU 层和 1 个 conv 层，将通道数为 3n 的特征图融合成通道数为 3 的特征图，上述操作将不同通道进行融合，提高了非线性表示能力，不仅得到了更好的增强图像，也达到了加快下一阶段训练速度的目的。图 B1.2 是原始微光图像和经过多尺度对数变换处理后的图像。

原始微光图像　　多尺度对数变换后的图像

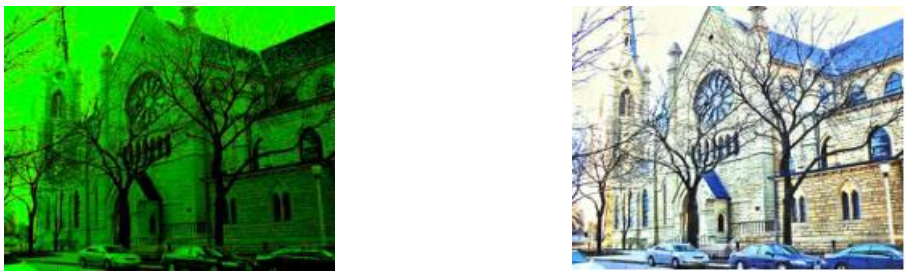
图 B1.2 原始微光图像和经过多尺度对数变换处理后的图像

3.2 卷积差分

卷积差分以多尺度对数变换处理后的结果 X1 为输入，输出结果 X2 为相同大小的形状。首先将 X1 经过多个卷积层进行特征提取，接着把每个卷积层输出的特征图利用一个 concat 层在通道维度上进行特征融合，最后通过一个 1×1 卷积核将融合后的特征图映射为光照分量。与 MSR 类似，得到光照分量后，用多尺度对数转换阶段输出的 X1 减去光照分量，得到反射分量 X2。

3.3 颜色恢复

MSR-net 是直接在 RGB 空间中进行处理，由于 RGB 各通道之间存在很强的色彩相关性，很容易出现色彩失真的现象。在论文 *A multiscale retinex for bridging the gap between color images and the human observation of scenes* 中提出了色彩恢复算法来解决这个问题。MSR-net 基于这个思路，在第三阶段中通过 1×1 的卷积进行色彩恢复。上述操作很好的解决了图像的色彩失真问题。图 B1.3 是颜色恢复之前的图像及颜色恢复之后的图像。



色彩恢复之前的图像　　色彩恢复之后的图像

图 B1.3 色彩恢复之前的图像及色彩恢复之后的图像

4 数据集

案例使用的数据集是 2018 年 Wei 等人构建的低光配对数据集(LOL)，包括两个类别：真实摄影对和原始图像合成对，前者捕获了真实情况下的正常光图像，后者是通过将正常光

图像从 RGB 转换到 YCbCr 通道，计算 Y 通道的直方图，调整 Y 通道的直方图使其与弱光图像相吻合，得到对应的低照度图像。数据集中共包含 500 对低/正常光图像，尺寸均为 400×600 的图形格式。部分数据集如图 B1.4 所示。

图 B1.4 LOL 数据集

5 要求

5.1 基本要求

⑴ 构建项目工程，正确配置源代码运行环境，并成功运行源代码；

⑵ 使用 LOL 数据集，并输出图像增强效果图。

⑶ 实验结果分析：选取 LOL 数据集中不同环境下的图像增强结果进行对比，并分析实验结果。

5.2 拓展要求

⑴ 采用不同于 LOL 数据集中的待增强图像进行测试，并分析实验结果。

⑵ 本案例算法虽能较好的解决色彩失真问题，但抑制噪声能力偏差，尝试更改网络结构，达到更好的增强效果。

5.3 撰写项目文档

包括 MSR-net 图像增强原理与流程、源代码说明文档、实验过程及结果分析等。

6 参考文献

[1] Shen L, Yue Z, Feng F, et al. Msr-net: Low-light image enhancement using deep convolutional network[J]. arXiv preprint arXiv:1711.02488, 2017.

B2 基于 KinD 算法的图像增强网络

本案例以 2019 年 zhang 等人发表的论文 Kindling the Darkness:A Practical Low-light Image Enhancer 为基础，介绍了基于 KinD 算法实现图像中增强的基本原理和实现流程，对核心源码进行了解析，并指导学生完成图像增强实验、分析实验结果。通过本案例的学习，可以扩展视野，了解深度学习在图像增强领域的前沿成果；通过阅读和学习优秀学术论文，培养学生自主学习的能力、工程实践能力和分析总结能力，提升科研兴趣和创新意识。

1 学习目标

⑴ 了解卷积神经网络在图像增强方面的应用；

⑵ 阅读文献、调试运行开源程序、分析对比实验结果，培养自主学习、研究的能力。

2 案例背景

图像是人类获取、表达和传递信息最重要的载体。然而图像在采集的过程中，常常会由于天气、光照强度、光源方向等多种复杂因素的干扰，导致图像出现对比度差、亮度过低或过高、细节丢失、色彩失真、含有大量噪声等问题。

图像增强是一种增强图像中有用信息的技术，目的是改善图像的视觉效果，便于图像适用到不同的场合。为了达到上述效果，众多学者进行了大量研究。

随着深度学习的不断发展，基于 CNN 的图像增强方法陆续出现，2019 年 zhang[1]等人提出了 KinD 增强算法，在去除噪声的同时还可以很好保留细节信息，增强后的图像也更符合实际情况。该算法以 Retinex-Net 为基础，使用不同光/曝光条件下捕获的成对图像进行训练，采用分而治之的思想对子网络进行了优化，在光照调解网络中还可以根据用户的不同需求灵活地调整光线等级。本案例以此论文为基础，介绍基于 KinD 的图像增强的算法原理、方法和 TensorFlow 实现，并分析、对比实验结果。

3 基于 KinD 的图像增强算法原理

整个网络[1]采用分而治之的思想，有两个分支分别对应于反射分量和光照分量。网络结构图如 B2.1 所示，从功能的角度来看，它可以分为：图像分解网络、反射分量恢复网络和光照分量调节网络三个模块。

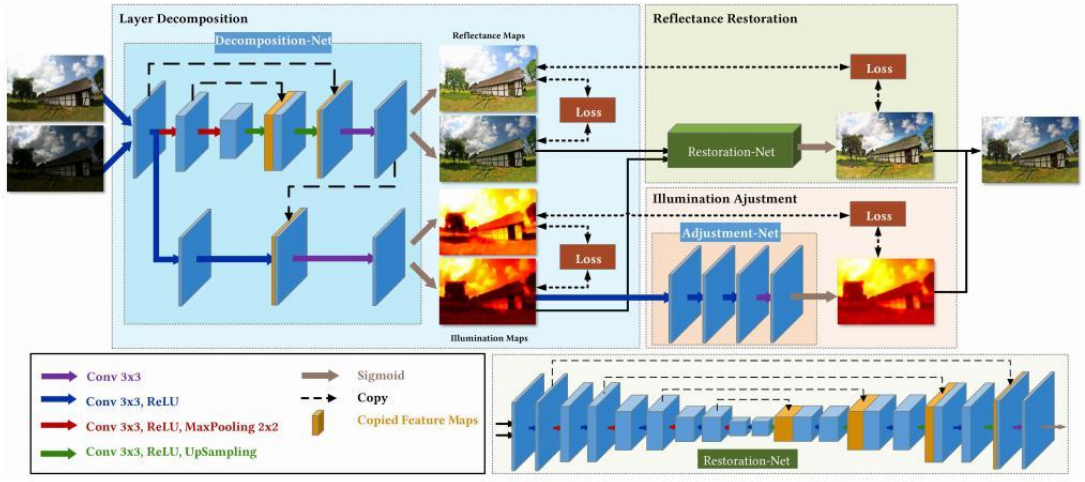


图 B2.1 网络结构图

3.1 图像分解网络

基于 Retinex[2]理论，该论文中也是将原图像分解为光照分量和反射分量，进而对这两个分量分别进行处理，再将处理后的两个分量进行相加得到最终增强图像。图像分解网络包含两个分支，分别对应于反射分量提取分支和光照分量提取分支。反射分量提取分支采用了典型的 5 层 U-Net，结构简单，在对应的采样层之间进行跳跃连接，在上采样恢复过程中会将当前层的抽象特征与浅层特征融合，更有效的提取特征。光照分量分支先使用两个 conv+ReLU 层处理后，和反射分量分支最后阶段的特征图进行拼接融合，此操作可以筛选和重组特征图中不同通道上的信息，确保网络的拟合能力。

3.2 反射分量恢复网络

论文中[1]认为低光图像中反射分量的噪声与光照分量强相关，因此，该部分将低光图像的光照分量同反射分量一起引入恢复网络进行训练，这样能够更好地去除噪声，同时尽可能保留细节信息。该部分的网络结构类似于图像分解网络中反射分量提取的分支，以 U-Net 网络为基础，但更深。实验结果如图 B2.2 所示(上为原图，下为恢复后图像)，黑暗区域的噪声很大程度上得到去除，而窗口区域的纹理（例如灰尘/水渍）可以很好地保留。此外还很好了避免了色彩失真问题。

图 B2.2 反射分量恢复网络结果

3.3 光照分量调节网络

在光照分量调节网络中，将低光图像分解出的光照分量输入至网络中，将高光图像分解出的光照分量作为光照调节网络的参考图像，共同训练光照分量调节网络，这样的设定可以使处理后的光照分量在相对黑暗的区域增加的光更少，而在明亮的区域增加的光更多或大致相同，使其更符合实际情况。该网络是轻量级的，包含 3 个层(两个 conv+ReLu，一个 conv

层)和 1 个 Sigmoid 层。

4 数据集

案例使用的数据集是 2018 年 Wei[2]等人构建的低光配对数据集(LOL)，包括两个类别：真实摄影对和原始图像合成对，前者捕获了真实情况下的正常光图像，后者是通过将正常光图像从 RGB 转换到 YCbCr 通道，计算 Y 通道的直方图，调整 Y 通道的直方图使其与弱光图像相吻合，得到对应的低照度图像。数据集中共包含 500 对低/正常光图像，尺寸均为 400×600 的图形格式。部分数据集如图 B2.3 所示。

图 B2.3　LOL 数据集

5 要求

5.1 基本要求

⑴ 构建项目工程，正确配置源代码运行环境，并成功运行源代码；

⑵ 使用 LOL 数据集完成模型训练和测试，输出图像增强结果图。

⑶ 实验结果分析：选取 LOL 数据集中不同场景下的图像增强结果进行对比，并分析实验结果。

5.2 拓展要求

⑴ 采用不同于 LOL 数据集中的低质量图像进行测试(例如 VV 数据集、MEF 数据集以及 DICM 数据集等)，并分析实验结果。

5.3 撰写项目文档

包括基于深度学习的图像增强算法的原理与流程、源代码说明文档、实验过程及结果分析等。

6 参考文献

[1] Zhang Y, Zhang J, Guo X . Kindling the Darkness: A Practical Low-light Image Enhancer[C]// 2019.

[2] Wei C, Wang W, Yang W, et al. Deep Retinex Decomposition for Low-Light Enhancement[J]. 2018.

B3 基于全局单应性的自由视角图像拼接网络

本案例以北京交通大学博士研究生聂浪[1]的论文 A View-Free Image Stitching Network Based on Global Homography 为基础，介绍了一种自由视角下图像拼接网络的基本原理和实现流程，对核心源码进行了解析，并指导学生完成图像拼接实验、分析实验结果。通过本案例的学习，可以拓展视野，了解图像拼接和深度学习领域前沿成果；通过阅读和复现优秀学术论文，培养学生自主学习的能力、工程实践能力和分析总结能力，提升科研兴趣和创新意识。

1 学习目标

⑴ 理解图像拼接网络的基本原理；

⑵ 理解深度学习实现全局单应性变换的原理和方法；

⑶ 阅读文献、调试运行开源程序，培养自主学习、研究的能力。

2 案例背景

图像拼接是指将具有空间重叠区域的两幅或多幅图像组合成全景图像，涉及到计算机视觉中图像处理、图像配准、图像融合等方面的知识。近年来，随着图像拼接技术研究层面的不断创新，其应用层面也在不断的扩展，得到了广泛关注。在医疗领域中，当医生使用超声波设备对患者病变部位进行检测时，超声波设备采集到的只是局部的视觉影像，采用图像拼接技术能够将患者不同位置的影像拼接成完整全面的多视角图，使得医生对病灶区域有了更为全面深入的了解和判断。在航空领域中，图像拼接技术可以对航拍图像和卫星遥感图像进行拼接整合，得到高质量、高分辨率的全景图像，满足对大体积物体细节观察的需要。在金融、城市建设、农业等多个领域中，图像拼接都有着各种应用。

深度学习能够充分提取图像特征，泛化能力强，已广泛应用于计算机视觉领域，在行人检测、对象分类、图像分割等领域都有优异表现。2020 年，北京交通大学博士研究生聂浪在 Journal of Visual Communication and Image Representation 上发表论文 A View-Free Image Stitching Network Based on Global Homography，提出了一种基于深度学习的图像拼接网络，能够完成对任意视角图像的拼接。本案例以此论文为基础，介绍了这篇论文中自由视角下图像拼接网络的实现过程。

3 基于全局单应性的自由视角图像拼接算法原理

原论文中图像拼接的方法分为三个阶段实现，分别是：单应性估计、结构拼接和内容修正。

⑴ 单应性估计

首先使用 4 个卷积模块进行特征提取，这里的每个卷积模块包括两个卷积层和一个最大池化层，进行 L2 归一化后采用全局相关层来学习两个特征映射之间的特征全局相似性，最后使用由 3 个卷积层和两个全连接层组成的回归网络来处理预测偏移量 f，采用直线线性变换算法将预测的偏移量 f 转换为对应的单应性矩阵 H。

⑵ 结构拼接

结构拼接阶段是基于空间变换网络。对于 ImageA，使用单位矩阵，变换为 IAW；对于 ImageB，使用单应性估计阶段得到的 H，变换为 IBW，IAW 和 IBW 通过平均融合得到结构

拼接结果。

⑶ 内容修正

单应性估计是由图像四个顶点的偏移量计算得来，不会将每个像素对齐，轻微的预测误差就会导致整个拼接结果视觉上的模糊。因此，在结构拼接阶段得到的是一个粗对齐拼接结果。内容修正阶段是一个 UNet 网络，输入是粗对齐拼接结果，输出是精确对齐拼接结果。

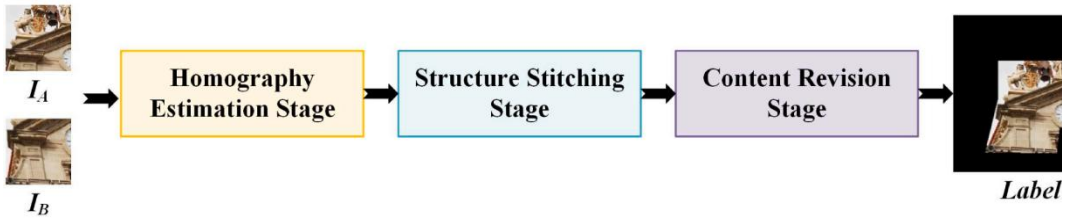


图 B3.1 基于全局单应性的自由视角图像拼接流程

4 数据集

训练神经网络通常需要大量的数据。然而，现有的图像拼接数据集只有少量的数据，这不足以用于深度学习方法。为了满足这一要求,原论文提出了一种可以生成大量用于图像拼接数据集的方法，本案例使用此方法并且利用 Microsoft COCO 2014Val 数据集来制作用于图像拼接的数据集。

Microsoft COCO 数据集的全称是 Microsoft Common Objects in Context，是一个评估计算机视觉模型性能的“黄金”标准基准数据集，旨在推动目标检测、实例分割、看图说话和人物关键点方面的研究。Microsoft COCO 数据集中的图像收集于复杂的日常生活场景，与 PASCAL、ImageNet 和 SUN 相比，其统计分析做得更为详尽。Microsoft COCO 数据集包含很多的分支，主要包括 2014Train/Val、2014Testing、2015Testing、2017Train/Val/Test、2017Unlabeled 等等。

图 B3.2 原论文数据集制作过程示意图

5 要求

5.1 基本要求

⑴ 构建项目工程，正确配置源代码运行环境，并成功运行源代码；

⑵ 实验结果分析：在不同环境、视差等场景下的图像拼接结果进行对比，并分析实验结果。

5.2 撰写项目文档

项目文档包括基于深度学习的图像拼接原理与流程、源代码说明文档、实验过程及结果分析等。

6 参考文献

[1] Nie L, Lin C, Liao K, et al. A view-free image stitching network based on global homography[J]. Journal of Visual Communication and Image Representation, 2020, 73: 102950.

B4 基于无监督深度学习的图像拼接

本案例以北京交通大学博士研究生聂浪[1]的论文 Unsupervised Deep Image Stitching: Reconstructing Stitched Features to Images 为基础，介绍了基于无监督深度学习图像拼接的基本原理和实现流程，讲解并指导学生完成图像拼接实验、分析实验结果。通过本案例的学习，学生可以拓展视野，了解图像拼接和深度学习领域前沿成果；通过阅读和复现优秀学术论文，可以培养学生自主学习的能力、工程实践能力和分析总结能力，提升科研兴趣和创新意识。

1 学习目标

⑴ 理解无监督深度学习的原理及其在图像拼接领域中的应用；

⑵ 阅读文献、调试运行开源程序、分析对比实验结果，培养自主学习、研究的能力。

2 案例背景

图像拼接是将两张或多张有一定重叠区域的图像进行拼接缝合，继而获得更广的视角，是计算机视觉中一项重要的研究内容。

在对实际图像场景进行拼接时，传统基于特征的图像拼接方法容易受到重叠区域、视角差异、实时环境情况（如光照强度、天气状况等）、图像场景中建筑物或物体本身属性（如建筑物的线条情况、物体的物理纹理情况等）的影响，从而导致此方法拼接效果往往不佳。为了解决上述问题，众多学者进行了大量研究，虽涌现出了许多不同且性能较好的图像拼接方法，但其对于各种真实环境的鲁棒性却依旧不佳。随着深度学习的不断发展，众多学者又将目光投向了将深度学习，希冀使用深度学习的方法来解决传统特征法无法解决的图像拼接问题。

2021 年，北京交通大学博士研究生聂浪在 IEEE Transactions on Image Processing 上发表论文 Unsupervised Deep Image Stitching: Reconstructing Stitched Features to Images，提出了首个基于真实场景的无监督深度学习图像拼接框架，本案例以此论文为基础，介绍基于无监督深度学习的图像拼接的算法原理、方法，并分析、对比实验结果。

3 基于无监督深度学习的图像拼接算法原理

原论文中基于无监督深度学习的图像拼接流程共分为两个阶段——无监督图像粗对齐阶段和无监督图像重建阶段，如图 B4.1 所示。其中左侧为无监督图像粗对齐阶段，右侧为无监督图像重建阶段。

在第一阶段的无监督学习图像粗对齐中，使用了一个基于消融损失来约束一个无监督单应网络，同时引入了一个拼接域变换层减小图像大小。在此阶段的开始，输入数据为两个高分辨率图像，先使用一个卷积网络对其进行了特征的提取，然后将提取到的特征结果作为输入，传递给之后的无监督单应网络用以估计图像的单应性，从而扭曲图像并进行粗对齐。之后将上一步的输出图像投入到拼接域变换层减少图像的无用部分（多余的黑色像素），从而降低图像大小和图像占用的空间，减轻第二阶段运算压力。在第二阶段的无监督学习图像重建中，使用了一个无监督图像重建网络用以消除特征到像素的重影。重建网络由两个分支实现：低分辨率变形分支（图 B4.1 右侧顶部）和高分辨率细化分支（图 B4.1 右侧底部），分别学习图像拼接的变形规则和提高分辨率。在低分辨率分支中，首先将第一阶段的输出进行下采样，然后通过编码器-解码器网络进行跳跃连接来学习如何变形图像，进而输出图像的内容掩码和接缝掩码。内容掩码用于约束重建图像的特征接近扭曲图像，而接缝掩码旨在约

束重叠区域的边缘自然和连续。之后，在高分辨率分支中利用三个独立的卷积层和八个资源块用以计算在此分支下的内容损失和接缝损失，再结合低分辨率中的内容掩码和接缝掩码细化拼接图像。最后输出两幅图像的拼接结果。

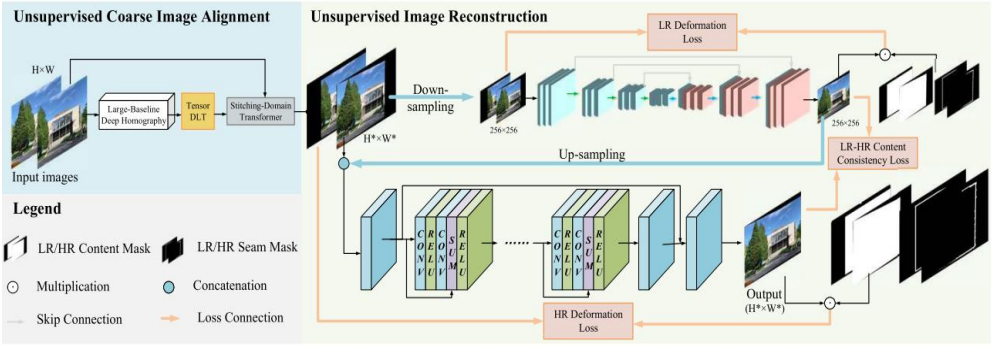


图 B4.1 基于无监督深度学习的图像拼接流程

4 数据集

案例所使用的数据集 UDIS-D 同样来自论文 Unsupervised Deep Image Stitching: Reconstructing Stitched Features to Images。该数据集为首个真实场景的无监督图像拼接数据集，其中包含了不同场景，不同重叠率和不同的视差的图片对。训练集共 10,440 对图像，测试集包含 1,106 对图像。部分数据如图 B4.2-图 B4.4 所示。

图 B4.2 不同场景

图 B4.3 不同重叠率

图 B4.4 不同视差

原数据集下载地址：https://pan.baidu.com/s/1bagitzF-MjIp1CN3DImHVg?pwd=1234

链接密码：1234

5 要求

5.1 基本要求

⑴ 构建项目工程，正确配置源代码运行环境，并成功运行源代码；

⑵ 使用 UDIS-D 数据集完成模型训练和测试，输出图像拼接结果图；

⑶ 实验结果分析：选取 UDIS-D 数据集中不同环境、视差等场景下的图像拼接结果进行对比，并分析实验结果。

5.2 拓展要求

⑴ 采用不同于 UDIS-D 数据集中的待拼接图像进行测试，并分析实验结果；

⑵ 分别实现基于特征匹配的传统图像拼接算法和案例 B4 所采用的基于深度学习的图像拼接算法，分析实验结果，并客观评价各种算法的拼接结果。

5.3 撰写项目文档

包括无监督深度学习图像拼接原理与流程、源代码说明文档、实验过程及结果分析等。

6 参考文献

[1] Nie L, Lin C, Liao K, et al. Unsupervised deep image stitching: Reconstructing stitched features to images[J]. IEEE Transactions on Image Processing, 2021, 30: 6184-6197.

B5 基于全卷积孪生网络的目标跟踪

本案例以牛津大学科学工程系团队[1]发表的论文 Fully-Convolutional Siamese Networks for Object Tracking 为基础，使用全卷积孪生网络 SiamFC 实现了目标跟踪。在本案例中，主要介绍了孪生神经网络的基本结构，对目标代码的结构和关键部分进行了解析，并提供了运行目标跟踪代码的完整步骤和实验结果。

通过本案例的学习，可以了解到孪生神经网络在目标跟踪领域的成果；通过阅读优秀外文文献，可以培养学生自主学习的能力、工程实践能力和分析解决问题的能力，提升科研兴趣和创新意识。

1 学习目标

⑴ 了解孪生神经网络的结构以及其在目标跟踪领域的应用。

⑵ 根据 SiamFC 论文，完成 SiamFC 目标跟踪模型。

⑶ 培养学生阅读文献，动手编程，分析问题并解决问题的能力。

2 案例背景

目标跟踪是计算机视觉的基础领域之一，也是一项具有挑战性的研究内容，受到众多学者的广泛关注。目标跟踪被广泛应用于民用和军用领域，如民用智能视频监控、汽车自动驾驶、智能人机交互和军事目标情报收集和精准打击、导弹制导等，无论在国防军事还是民用方面都具有重要的研究意义和广阔的应用前景。

近年来，由于卷积神经网络提取的深度特征鲁棒性好、描述能力强，在目标跟踪领域中渐渐取代了传统手工设计的特征。

孪生神经网络由两个分支的神经网络组成，而这两个分支的神经网络的权重是共享的，利用双分支的输出可以计算出两条分支的输入之间的相似度，因此最初用于银行系统的客户签名验证。 SINT 是第一个使用孪生神经网络用于目标跟踪的算法，将目标跟踪任务看作是一种相似度度量的问题，为目标跟踪提供了一种新思路。后来 Bertinetto 等人提出的全卷积孪生网络跟踪算法(SiamFC)，简化了相似度的计算过程，大大提高了跟踪速度，证明孪生跟踪算法在精度和速度上的巨大潜力。

本案例以 SiamFC 算法为主要内容，介绍基于全卷积孪生网络的目标跟踪算法原理、和 Pytorch 代码实现。

3 基于全卷积孪生网络的目标跟踪

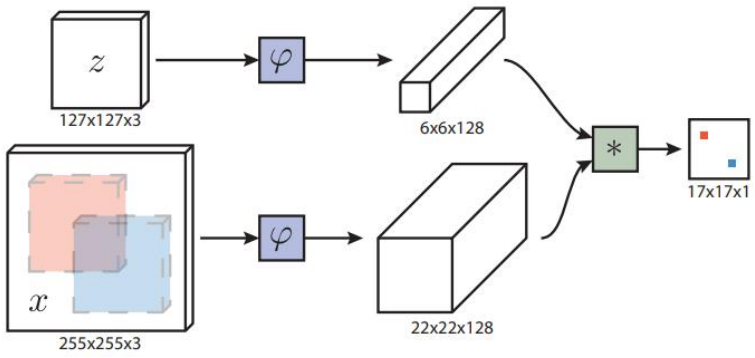
传统的单目标跟踪算法多为在线跟踪，在线更新模型。这种方法的好处就是速度快，但是跟踪质量并不是很高。

深度学习方法多为离线训练，在线跟踪，这样做的好处就是跟踪质量好，但是速度比不上相关滤波的方法，但是从 SiamFC 开始，基于深度学习的方法在速度上已经可以和传统的相关滤波并驾齐驱，甚至更优。主要原因是因为 Siam 系列基于孪生结构，简单高效，接下来我们详细介绍这一结构。

所谓孪生结构，具体来说就是该结构有两个输入，一个是作为基准的模板，另一个则是要选择的候选样本。

在单目标跟踪任务中，作为基准的模板则是我们要跟踪的对象，通常选取的是视频序列第一帧中的目标对象，而候选样本则是之后每一帧中的图像搜索区域，而孪生网络要做的就

是找到之后每一帧中与第一帧中的范本最相似的候选区域，即为这一帧中的目标，这样我们就可以实现对一个目标的跟踪。本案例使用的孪生网络结构如下：



B5.1 SiamFC 孪生网络结构

该结构首先 z 为输入的范本，即第一帧图像中的目标框，大小为 127x127x3，x 为输入的搜索图像，大小为 255x255x3，接着对两个输入分别进行φ变换（作者采用了 AlexNet 的网络结构），即特征提取，分别生成了 6x6x128 和 22x22x128 的特征图，提取了特征之后，再对提取的特征进行互相关操作（即求卷积），生成响应图，互相关操作如下：

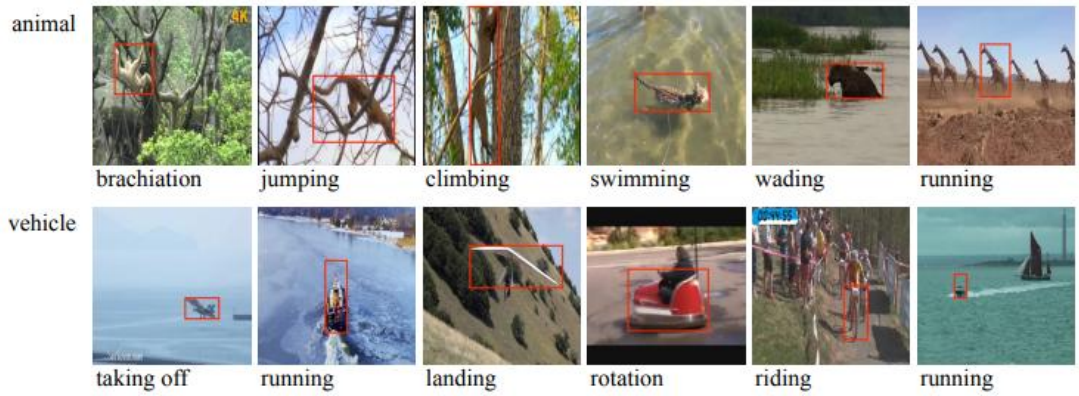
$$f(z,x) = \varphi(z) \star \varphi(x) + b\Pi \tag{1}$$

其中， bΠ 为每个位置对应的值，$\star$ 为卷积运算，通过卷积运算提取 x 中与 z 最为相近的部分.

卷积左边对应的是目标 z 的特征图，右边为搜索区域 x 的特征图，最终生成的是响应图，响应值最高的位置就对应着 z 可能的位置。

4 目标跟踪数据集 GOT-10k

GOT-10k 是一个大型的目标跟踪数据集，包含近 10000 个含有真实移动对象的视频片段和超过 150 万个手动标记的边界框。



B5.2 GOT-10K 数据集

在训练集中，包含 9335 个视频序列，563 种不同的目标类别以及 83 种不同的运动类型，比如滑雪，跑步等。在验证集和测试集中，分别包含 180 个视频序列，84 种目标类别以及 32 种运动类型。

在 GOT-10K 官网可以下载到完整的数据集。

GOT-10K: http://got-10k.aitestunion.com/index

```
|-- GOT-10k/
    |-- train/
    |   |-- GOT-10k_Train_000001/
    |   |    ......
    |   |-- GOT-10k_Train_009335/
    |   |-- list.txt
    |-- val/
    |   |-- GOT-10k_Val_000001/
    |   |    ......
    |   |-- GOT-10k_Val_000180/
    |   |-- list.txt
    |-- test/
    |   |-- GOT-10k_Test_000001/
    |   |    ......
    |   |-- GOT-10k_Test_000180/
    |   |-- list.txt
```

B5.3 GOT-10K 数据集结构

除了包含视频序列以外，GOT-10K 数据集中的每个序列文件夹还包含 4 个注释文件和 1 个元文件。下面对这些文件的简要说明。N 为视频序列长度。

groundtruth.txt：一个 N×4 的矩阵，每一行表示对应帧中的对象位置。

cover.label：一个 N×1 数组，表示对象被遮挡的比率。

absense.label：一个 N×1 二进制数组，表示每帧中是否存在对象。

cut_by_image.label：一个 N×1 二进制数组，表示每帧中的对象是否按图像裁切。

meta_info.ini：有关序列的元信息，包括对象和运动类、视频 URL 等。

```
[METAINFO]
url: https://youtu.be/A7COfm1iCUE
begin: 00:00:20
end: 00:00:30
anno_fps: 10Hz
object_class: duck
motion_class: swimming
major_class: bird
root_class: animal
motion_adverb: slowly
resolution: (1280, 720)
```

B5.4 meta_info.ini

5 要求

5.1 基本要求

⑴ 构建 SiamFC 目标跟踪网络模型，搭建运行环境。

⑵ 使用 GOT-10K 数据集完成模型训练和测试。

⑶ 运行 SiamFC 模型，并实现目标跟踪。

5.2 拓展要求

⑴ 在不同目标跟踪数据集上对 SiamFC 进行测试，并观察对比跟踪效果。

⑵ 查阅相关资料，了解其他在 SiamFC 算法上进行改进的其他目标算法，运行相关代码，并进行测试评估。

5.3 撰写项目文档

包括基于全卷积孪生网络的目标跟踪案例的简介、原理与方法，程序流程与源码解析、实验过程及结果分析等。

6 参考文献

[1] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking[C]//European conference on computer vision. Springer, Cham, 2016: 850-865.

B6 基于 VGG19 的图像风格迁移

图像风格迁移是指给定一张普通图片和一张艺术风格图片，在保留目标图片内容的基础上，将风格图片应用于目标图片上，生成一张呈现艺术风格和普通图片内容的迁移图片。本案例使用 VGG19[1]实现了图像风格迁移。

通过本案例的学习，可以了解到图像风格迁移的原理；通过阅读优秀外文文献，可以培养学生自主学习的能力、工程实践能力和分析解决问题的能力，提升科研兴趣和创新意识。

1 学习目标

⑴ 理解图像风格迁移的原理；

⑵ 理解 VGG19 网络架构，并能够应用它进行特征提取；

⑶ 能够根据指定任务，自主查阅资料，编写和调试程序，并分析实验结果。

2 案例背景

早期的图像风格迁移技术算法适用风格范围窄且迁移转换结果不理想。随着人工智能和深度学习的兴起，基于深度学习的图像风格迁移技术快速发展使得该技术被广泛的应用于图片影像加工美化。目前很多智能手机 APP 中都应用了图像风格迁移技术，例如美图工具中的各类滤镜。风格本质上是指在各种空间尺度上图像中的纹理，颜色和视觉图案。图像风格迁移是指给定一张普通图片和一种艺术风格图片，在保留目标图片内容的基础上，将图片风格应用在目标图片上，生成一张呈现艺术风格和普通图片内容的迁移图片。

本案例以实现图像风格迁移为主要内容，介绍基于 VGG19 的图像风格迁移原理和代码实现。

3 图像风格迁移原理

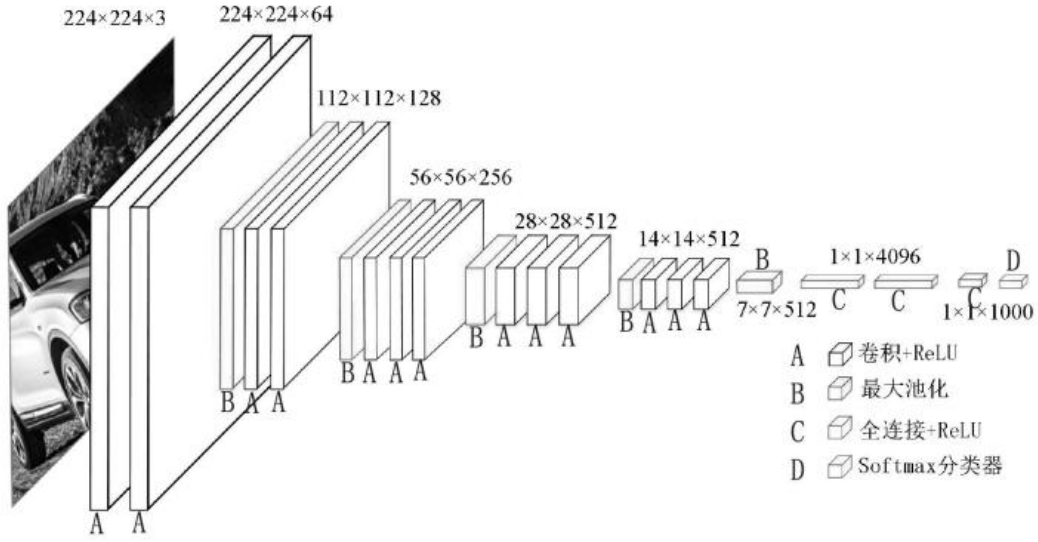


图 B6.1 VGG19 结构图

VGG19 在图像风格迁移中提取特征时，输入特征后直接输出对应这种特征的图片。卷积的过程就是特征提取的过程，每一个卷积核代表一种特征。在可视化卷积核的过程中，低层次的卷积核对目标图像的颜色、边缘信息感兴趣，提取的特征是简单的检测点和线等。随着层次的不断加深，内容越来越抽象和复杂，卷积层感知图像中的物体位置更准确，提取的

特征通常是复杂的特定物体。

选择深层“block5_conv2”卷积层提取内容特征；选择“block1_conv1”至“block5_conv1”的五个从低维到高维过渡的卷积层提取样式特征。浅层特征表示图片的风格，深层特征表示图片的内容。图片的内容其实就是大致的轮廓，而图片的风格则是与色彩、纹理相关的细节。

图像风格迁移包括：图像内容获取；图像风格提取；内容和风格融合。

⑴ 首先，使用 VGG 中的一些层的输出来表示图片的内容特征和风格特征。

⑵ 将内容图片输入网络，计算内容图片在网络指定层上的输出值。

⑶ 计算内容损失。内容损失为内容图片在指定层上提取出的特征矩阵与噪声图片在对应层上的特征矩阵的差值的 L2 范数。L2 范数是指向量各元素的平方和然后求平方根。

对应每一层的内容损失函数：

$$L_i = \frac{1}{2 * M * N} \sum_{ij} (X_{ij} - P_{ij})^2$$

其中，X 是噪声图片的特征矩阵，P 是内容图片的特征矩阵。M 是 P 的长*宽，N 是信道数。最终的内容损失为，每一层的内容损失加权和，再对层数取平均。

⑷ 将风格图片输入网络，计算风格图片在网络指定层上的输出值。图片风格使用卷积层特征值的 Gram 矩阵表示。

⑸ 计算风格损失。风格损失为风格图像和噪声图像特征矩阵的 Gram 矩阵的差值的 L2 范数。

对于每一层的风格损失函数：

$$L_i = \frac{1}{4 * M^2 * N^2} \sum_{ij} (G_{ij} - A_{ij})^2$$

其中 M 是特征矩阵的长*宽，N 是特征矩阵的信道数。G 为噪音图像特征的 Gram 矩阵，A 为风格图片特征的 GRAM 矩阵。最终的风格损失为，每一层的风格损失加权和，再对层数取平均。

⑹ 将内容和风格相融合，最终用于训练的损失函数为内容损失和风格损失的加权和。

$$L_{total} = \alpha L_{content} + \beta L_{style}$$

⑺ 当训练开始时，根据内容图片和噪声，生成一张噪声图片。并将噪声图片喂给网络，计算 loss，再根据 loss 调整噪声图片。将调整后的图片喂给网络，重新计算 loss，再调整，再计算……直到达到指定迭代次数。此时，噪声图片已兼具内容图片的内容和风格图片的风格，进行保存即可。

4 要求

4.1 基本要求

⑴ 构建项目工程，正确配置源代码运行环境，并成功运行源代码；

⑵ 使用 VGG19 预训练模型完成图像风格迁移，输出迁移后的结果图；

⑶ 实验结果分析：选取不同风格，不同 epoch 的图像迁移结果进行对比，并分析实验结果。

4.2 拓展要求

⑴ 采用自己的数据集重新进行训练，并分析实验结果。

⑵ 分别使用 VGG16 与 VGG19 完成图像风格迁移，分析实验结果，并评价不同算法的迁移效果。

4.3 撰写项目文档

包括图像风格迁移原理与流程、源代码说明文档、实验过程及结果分析等。

5 参考文献

[1] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.