**图像处理与分析调研报告**

* + 通过查阅文献资料，列举并简要介绍图像处理与分析的若干应用方向。
  + 针对以上的任意一个应用方向展开调研，总结其国内外研究现状。如针对目标检测，已有哪些经典的检测方法，对经典的和目前流行的检测方法进行简要介绍。
  + 从选择的方向中，选择一篇文献（中英文均可），通过阅读该文献，理解其提出的算法。用自己的语言介绍其提出的方法，并进行分析。

提交截止日期：12月12日

报告电子版（件名统一为学号\_姓名）提交给课代表

[501班交给张永朋2084954551@qq.com](mailto:501班交给张永朋2084954551@qq.com)

[502班交给邹逸飞1156076229@qq.com](mailto:502班交给邹逸飞1156076229@qq.com)

其他同学交给张永朋

附：报告模板

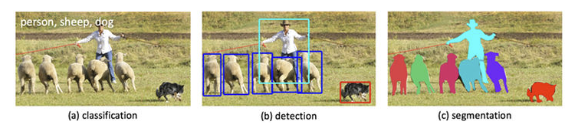
**图像处理与分析调研报告**

**学号: 162050127**

**姓名: 颜劭铭**

1. 图像处理与分析的若干应用方向简介

Low level的任务主要包括图像复原方面，去噪，去模糊，超分辨率，图像修复，低光增强等等；High Level的主要包括分类，分割，检测，图像理解等等。



1. 目标检测：

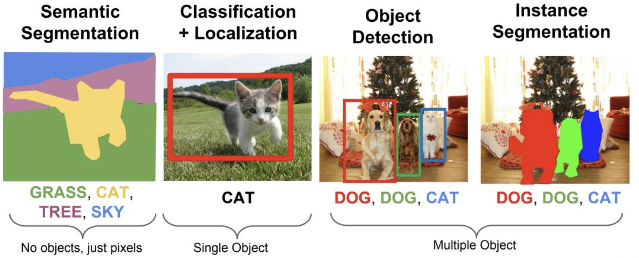
从图像的背景中分离出所有感兴趣的目标（物体），确定它们的类别和位置。基于深度学习的算法主要分为两类：Two Stage和One Stage。主要的应用包括人脸检测，行人检测，车辆检测，遥感监测等。

1. 分割：

图像分割是对图像中属于特定类别的像素进行分类的过程，因此图像分割可以认为是按像素进行分类的问题。

传统的图像分割算法均是基于灰度值的不连续和相似的性质。而基于深度学习的图像分割技术则是利用卷积神经网络，来理解图像中的每个像素所代表的真实世界物体。

基于深度学习的图像分割技术主要分为两类：语义分割及实例分割。



* 1. 语义分割：

为图像中的每个像素分配一个类别，但是同一类别之间的对象不会区分。

* 1. 实例分割

只对特定的物体进行分类。看起来与目标检测相似，不同的是目标检测输出目标的边界框和类别，实例分割输出的是目标的Mask和类别。

主要应用有人俩识别，抠图相机，虚拟化妆，在线试衣间，零售图像识别，手写字符识别，医学影像，自动驾驶等。

1. 图像分类

从给定的分类[集合](https://so.csdn.net/so/search?q=%E9%9B%86%E5%90%88&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/Avery123123/article/details/_blank)中给图像分配一个标签的任务。即分析一个输入图像并返回一个将图像分类的标签。标签来自预定义的可能类别集。

1. 图像处理
   1. 图像复原/图像重建：

图像复原是一个客观的过程，针对质量降低或失真的图像，试图恢复其原始的内容或质量。复原技术是面向退化模型的，并且采用相反的过程进行处理，以便恢复出原图像。在进行图像复原之前要先建立起其退化模型，根据该模型进行图像复原。

因此主要方法思路是由退化后的图像估计出退化函数和噪声函数，然后可以得到恢复算子。恢复计算，可以在空域上进行恢复，也可以在频域上进行恢复。

* + 1. 超分辨率：

是指利用光学及其相关光学知识，根据已知图像信息恢复图像细节和其他数据信息的过程，简单来说就是增大图像的分辨率，防止其图像质量下降。

方法主要分为传统方法和基于深度学习的方法，传统方法包括基于插值的方法，基于重构的方法和基于浅层学习的方法。其中基于浅层学习的方法主要包括机器学习、流形 学习、样本学习和稀疏编码等，用于数据量较小的情况，人为设计特征过程较复杂。

基于深度学习的方法可以分为基于卷积神经网络的SR方法、基于残差网络的SR方法和基于生成对抗网络的SR方法。

* + 1. 图像去噪/去雨/去雾/去模糊：

这类问题和超分辨率的问题一样都是属于图像复原问题，同样可以分为传统方法和基于深度学习的方法。

* 1. 低光增强：

低光增强是图像增强的一种，图像增强是一个主观的过程，其目的是改善图片的质量，对于感兴趣的部分加以增强，对于不管兴趣的部分予以抑制。而低光增强主要针对的是对于欠曝光的图片，将其恢复到正常曝光。

1. 所选择的研究问题的背景

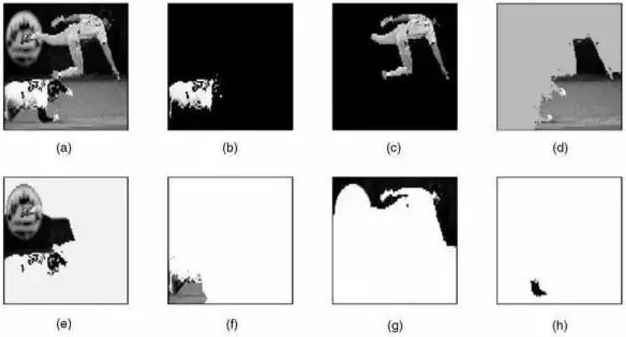
我选择的研究问题是语义分割。

语义分割是在像素级别上的分类，属于同一类的像素都要被归为一类，因此语义分割是从像素级别来理解图像的。

在深度学习火热之前，传统的语义分割主要指的是普通的图像分割，这个时期的分割由于计算机计算能力有限，早期只能处理一些灰度图，后来才能处理rgb图，并且主要是通过提取图片的低级特征，然后进行分割，涌现了一些方法：Ostu、FCM、分水岭、N-Cut等。这个阶段一般是非监督学习，分割出来的结果并没有语义的标注，换句话说，分割出来的东西并不知道是什么。

传统的图像分割方法主要有基于阈值的图像分割方法：通过设定不同的特征阀值，把图像像素点分为具有不同灰度级的目标区域和背景区域的若干类。它特别适用于目标和背景占据不同灰度级范围的图，目前在图像处理领域被广泛应用，其中阀值的选取是图像阀值分割中的关键技术；基于区域的分割方法：以直接寻找区域为基础的分割技术，具体算法有区域生长和区域分离与合并算法；基于边缘检测的方法：通过检测包含不同区域的边缘来解决分割问题；还有基于马尔科夫随机场，遗传算法，聚类等的图像分割方法。

传统方法进行语义分割有一个缺点：在实际运用中，每运行一次N-cut，只能切割一次图片，为了分割出图像上的多个物体，需要多次运行，下图展示了对原图a进行7次N-cut后，每次分割出的结果。缺点：效率低、准确度不高。



而当深度卷积网络流行之后，基于深度学习方法的语义分割效果比传统方法提升了很多。

1. 该问题的国内外研究现状

在这个部分主要总结的是基于深度学习方法中一些较为经典的语义分割模型。

FCN：[1]

1. FCN发布于2014年，是语义分割领域全卷积网络的开山之作

(2)其主要思路是将图像分类的网络改良成语义分割的网络，通过将分类器变成上采样层来恢复特征图的尺寸，进行端到端训练

(3)分类器变成上采样，这部分思想作者主要的解释是全连接层是一种特殊的卷积

(4)选择了AlexNet、GoogLeNet和VGG作为backbone，VGG效果最好，但是推理最慢

(5)最核心的思想是特征图的融合：假设最后的输出为pool5产生的x，利用转置卷积上采样，放大32倍，得到FCN-32s；将x上采样放大2倍，和pool4产生的特征图直接相加，再上采样放大16倍，得到FCN-16s；将FCN-16s进行上采样放大2倍，与pool3产生的特征图直接相加，在放大8倍，得到FCN-8s。在实验中，FCN-8s的效果最好

(6)backbone是分类网络，下采样都是maxpooling，上采样使用的是双线性插值初始化的转置卷积

U-Net：[2]

(1)U-Net发表于2015年，用于医学细胞分割。

(2)采用了编码器-解码器架构，通过四次下采样（maxpooling），四次上采样（转置卷积），形成了U型结构。这种U型结构，下采样的过程可以捕捉语义信息，与之对应的上采样过程可以进行精确地定位。这个网络可以利用非常少的图片进行端到端训练。

(3)使用了SGD+Momentum，损失函数为交叉熵。

(4)可以应对小样本的数据集进行较快、有效地分割，能够泛化到很多应用场景中去。

SegNet：[3]

(1)SegNet发布于2015年，使用编码器-解码器结构

(2)其backbone是2个VGG16，去掉全连接层（13层），对应形成编码器-解码器架构

(3)最核心的想法是提出了maxpool的索引来上采样的方法，从而免去了学习上采样的需要，在推理阶段节省了内存，因此更为高效

(4)作者说这个idea是来自于无监督特征学习。在解码器中重新使用编码器池化时的索引下标有这么几个优点：a. 能改善边缘的情况；b. 减少了模型的参数；c. 这种能容易就能整合到任何的编码器-解码器结构中，只需要稍稍改动

PSPNet：[4]

(1)PSPNet发布于2017年

(2)核心idea是提出了金字塔池化模块，模型带有空洞卷积

(3)金字塔池化融合了四个比例的特征，结合多尺寸信息SPP。最粗糙的1×1是全局尺度的池化，剩下的层次会将图像分为不同子区域，形成不同区域的信息表示。金字塔池模块中不同level的输出包含比例不同的feature map（比如输入的维度都是2048，有四个层次的金字塔，那么输出的维度则为2048/4=512）。为了保持全局特征的权重，若如果金字塔的数量为N，则在每个金字塔级别之后使用1×1卷积层将上下文表示的维度减小到原先的1/N。然后直接对feature map进行双线性插值，恢复到输入的长宽上。最后，将不同level的特征拼接起来作为金字塔池化的全局特征。文中给出的金字塔池化模块是一个四级模块，其大小分别为1×1、2×2、3×3、6×6。

(4)其backbone为修改Resnet-101为 ResNet-103，而且有辅助 loss，上采样是双线性插值

DeepLab v1：[5]

(1)发表于2014年

(2)核心思想是使用空洞卷积扩大感受野，条件随机场细化边界

(3)backbone是VGG16，下采样8倍

(4)利用深度卷积神经网络（DCNN）和概率图模型来进行语义分割，但目前的模型对于目标分割不足够精确。这是由于DCNN对于高层语义信息拥有很好的能力，但是对于低级的位置信息不足。本模型利用全连接条件随机场（fully-connected CRF）克服了这个问题。

DeepLab v2：[6]

(1)发表于2016年

(2)和v1的区别主要是在于多了空洞空间金字塔池化,在多个尺度上鲁棒地分割图像。ASPP使用多个采样率和有效视场的卷积核来检测传入的卷积特征，从而以多个尺度捕获目标和图像的上下文内容

(3)流程上是DCNN + Atrous convlution + CRF

(4)backbone是VGG16和ResNet-101，下采样8倍（ResNet需要在第二个stage第一个bottleneck中stride为2）

(5)结合DCNN和概率图模型的方法来改进对象边界的定位。DCNN中通常采用max-pooling和downsampling来实现不变性。我们通过将DCNN的输出和全连接条件随机场（CRF）结合来解决这个问题，在定性和定量上显示了CRF能够改善定位的性能。

DeepLab v3：[7]

(1)发表于2017年

(2)和v2的区别在于ASPP多了image-level feature，没有使用CRF

(3)提出了mutil-grid，通过调整膨胀系数可以来控制空洞卷积（级联或者并行的方式）捕捉多尺度的上下文信息，改进了级联网络的性能

(4)流程上是端到端训练了，backbone是ResNet-101

(5)下采样有8倍和16倍两种，8倍性能更好

(6)计算损失时，将输出的上采样，而非是将真实注释下采样

DeepLab v3+：[8]

(1)发表于2019年

(2)和v3的区别是多了一个解码器模块，backbone用了Aligned Xception（其中有深度可分解卷积）

(3)在DeepLab v3的基础了添加了一个简单有效地的解码器模块来校正分割的结果，尤其是目标的边界。

1. 所阅读的文献采用的方法及分析

此次阅读的文献是一篇CVPR2022的ORAL，题目为Semantic-Aware Domain Generalized Segmentation[9]。

摘要：

本文分别基于Instance Normalization (IN)与Instance Whitening (IW) 提出了两个用于编码器与解码器之间的即插即用模块：Semantic-Aware Normalization (SAN)与Semantic-Aware Whitening (SAW)，能够极大的提示模型的泛化能力。在面临各种与训练数据的分布不一致的测试数据时，SAN与SAW仍能帮助模型尽可能的维持模型的性能，可以说是竞赛乃至现实应用中的一大提点利器。

动机及背景知识：

在语义分割的无监督学习中，数据可知但是标签数据不可知，而且在现实生活中，我们无法得知包括图像数据在内的目标域任何信息，也就是说，模型在测试时可能会面临各种各样数据分布的图片，如果模型的[泛化](https://so.csdn.net/so/search?q=%E6%B3%9B%E5%8C%96&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/m0_61899108/article/details/_blank)能力不够，那么其性能肯定会出现很大的波动。

增强模型的泛化能力一个最常用的方式就是数据增强，即把训练数据转换为各种各样的形式使得模型在训练阶段就见过各种各样的数据分布，从而提高模型的泛化能力。但是如果希望数据增强能够使得转换后的训练数据覆盖所有测试数据的分布是不现实的。因此，通过数据增强来增强模型的泛化能力具有其固有的缺陷。

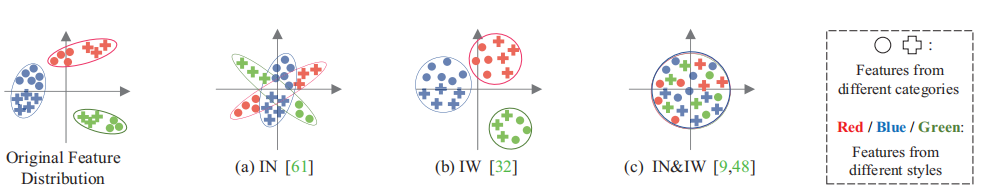


图1 应用不同的方法后，同一模型的编码器所提取的来自不同数据分布（不同域）的测试图片的特征分布。

另一个增强泛化能力的方式是使用Normalization 和 Whitening，该方法利用实例归一化(Instance Normalization,IN)或实例白化(Instance Whitening，IW)对不同样本的特征分布进行标准化。IN分别对单个图像的每个通道的特征进行标准化归一化，以减轻由于样式变化引起的特征不匹配。

从图1(b)也可观察得到，特征虽然均匀分布了，但是却也没有对齐特征的联合分布。最近有论文研究表明，如图1(c)所示，联合IN与IW后，能够对齐来自不同域（即数据分布不同）的联合特征边缘分布。然而，图1(c)中也可以观察到虽然特征边缘分布得到了对齐，但是条件分布却依然处于没有对齐的状态，每个类别的分布仍然混合在一起以至于难以区分。

由此引出了本文的出发点：既要对齐特征的全局边缘分布，也要对齐条件分布，从而使得每个类别的分布在特征空间能够被很好的区分开。具体做法便在于在IN与IW的基础上引入了类别信息。

作者为了解决语义分割模型的领域泛化问题，即使得训练不需要任何目标域数据，由于现有的方法在类与类之间的决策边界不明显，因此作者在分割中引入了更多的显式类信息，支撑分割模型。

方法：

整体方法框架如图2所示，具体来说，就是在原有的分割模型编码器-解码器结构框架中插入了两个即插即用的模块：Semantic-Aware Normalization (SAN)与Semantic-Aware Whitening (SAW)。它们分别基于IN与IW而设计的。

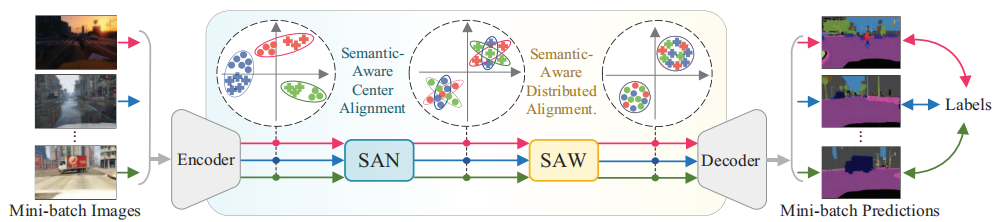


图2 模型总体框架图

SAN的总体框架如图3所示。

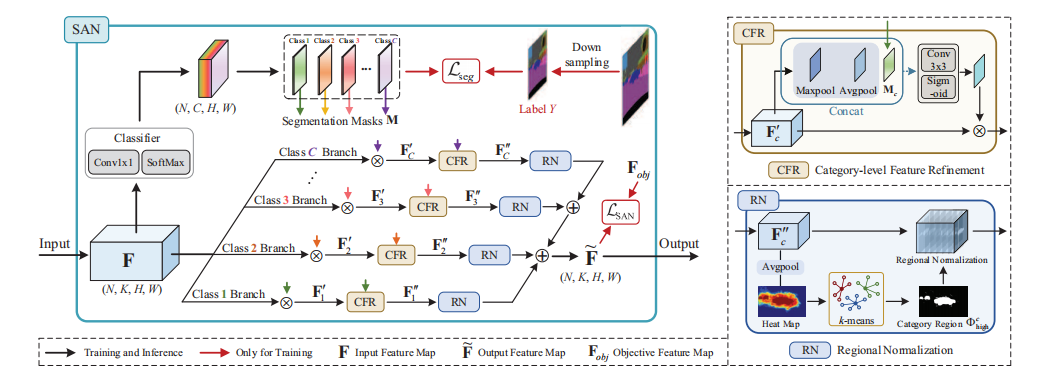


图3 SAN总体框架图

SAN模块主要是为了对每一类目标特征，分别进行特征归一化。

其中CFR模块本质上一个注意力模块，RN模块是一个区域归一化模块，首先通过对每类目标特征平均池化进行特征平滑，再借助某一类的不同目标实例进行聚类分割，最后根据分割区域对于区域内特征进行归一化。

首先对于编码器提取的特征图F提取出来每一类，并利用二值掩膜和特征相乘，得到每一类的单独特征，经过CFR注意力筛选重要区域，然后利用RN模块，先用聚类实现每类不同实例分割，然后对于分割区域进行区域级的特征归一化。

SAW的通体框架如图4所示。

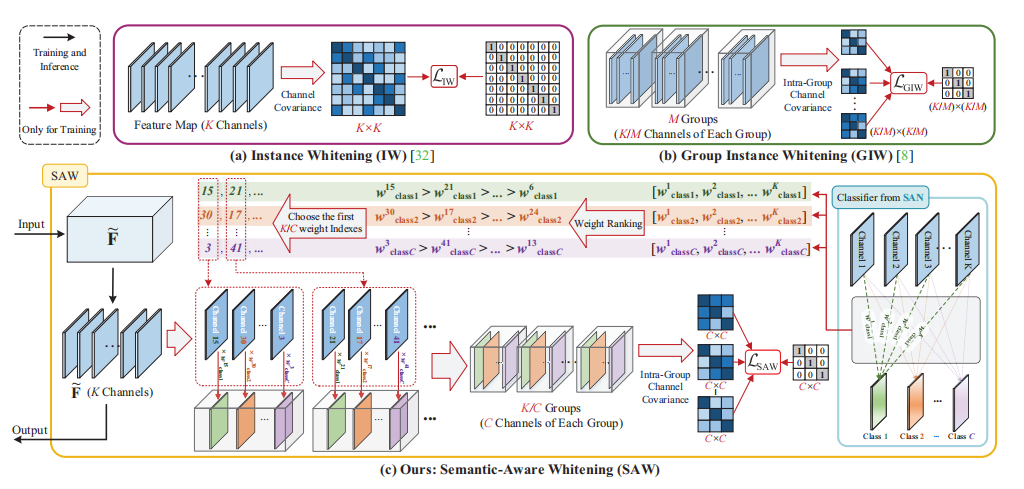


图4 SAW总体框架图

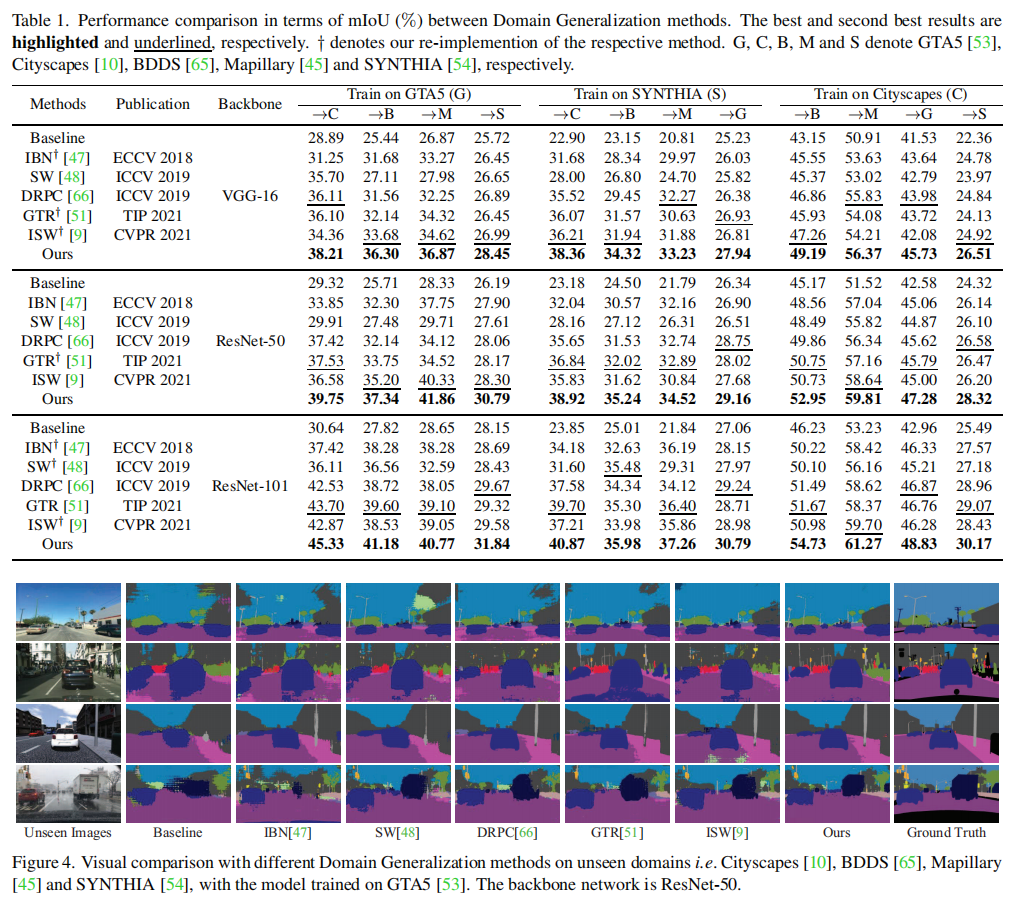
Whitening即让特征输出标准化。

IW：最小化输出特征通道间协方差，也就是让特征通道间相互关联最小，每一特征都有独一无二的用武之地；GIW：GIW在IW的基础上，将特征通道分组，然后在组内实施IW。SAW：SAW本质上属于GIW，只是它的分组不是简单的街区分组，而是将对每一类具有较大影响的通道分到一组。也就是说IW就是特征通道间解耦，GIW为组内特征通道解耦，SAW是一种特殊的GIW。

具体来说，SAW对于输入的K各通道选取了每个类权重参数最大的特征通道，即C个特征通道，然后对于特征通道进行分组，每组包含C个通道，每个通道都是对相应的语义类影响前列的通道，实现了一种特殊的分组，得到了K/C个组，然后实行GIW，实现了组内特征通道解耦，得到了每一类在k个特征通道的分类参数。

实验效果：

G,C,B,M分别代表GTA5,CityScapes，BDDS，Mapillary和SYNTHIA数据集，可以看出作者在各个数据集上的域泛化效果都达到了SOTA。



1. 学习图像处理与分析课程的收获

在学习了一学期的图像处理与分析课程后，首先我了解到了许多相关知识，如灰度变换与空间滤波，频率域滤波，图像复原与重建，彩色图像处理，形态学图像处理，图像分割等，并了解了相关方法。这些方法主要是传统方法，比如说中值滤波，直方图均衡化等等。

在这个深度学习盛行的时代，我们不可否认深度神经网络在效果上比大多数传统方法要来的好，但是这是建立在庞大的算力的基础上。我们不能够放弃对于传统方法的学习，很多传统方法在深度学习的数据预处理的时候也是必须要使用的，而很多时候，传统方法背后隐含的思想对于深度学习无论是网络的结构改进，或者是对于数据的预处理，总有意想不到的效果。特别是在自己实践对于一些图像进行处理后，可以发现处理的效果是还可以的，并且不需要太大的算力。

总的来说，通过课程学习，我虽然只掌握了图像处理技术的皮毛，但是对于很多方法有了进一步的了解，更加理解了一些处理方法的本质，包括图像变换、图像增强、图像复原、图像重建等等，这些传统方法的知识对于后续我在计算机视觉方向无论是低层次或者是高层次的信息的处理方法的学习将会有着很大的帮助，比如说先用传统方法处理，然后作为深度学习框架的输入使用；或者是先用深度学习网络学习源数据的特征表示，然后作为传统图像算法的输入；还有传统和深度学习方法并行处理，最后设计一个新的损失函数来结合使用等等。

1. 参考文献列表

[1]Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[2]Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[3]Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.

[4]Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.

[5]Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv preprint arXiv:1412.7062, 2014.

[6]Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.

[7]Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.

[8]Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.

[9]Peng D, Lei Y, Hayat M, et al. Semantic-aware domain generalized segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 2594-2605.