

项目选题

大部分图像处理任务可以理解为求解形如 $Au + \epsilon = f$ 的反问题。其中 u 代表原始数据， f 代表我们的观测， A 对应于线性算子，例如去噪中对应恒等映射，去模糊中对应卷积算子，CT图像中对应Radon变换等。 ϵ 对应于噪音。

传统图像处理方法往往对应于求解 $u^* = \arg \min_u \|Au - f\|_2^2 + \lambda R(u)$ ，参考项目1。深度学习方法试图“学”出 $f \rightarrow u$ 的映射。有些方法通过神经网络直接逼近反问题的映射 $u^* = \mathcal{F}_\theta(f)$ ，参考项目2；另一些借用自编码器的思想，将图像表示为对于编码 z 的解码 $u^* = \mathcal{U}_\theta(z)$ ，对应于项目3。对于这些不同的方向，从三个选题中任选一个你(们)喜欢的进行探索。

1 传统方法中的patch-based模型

对于图像 $u \in \mathbb{R}^{n \times n}$ ，令 P_j 是投影算子，将图像映射到第 j 个 patch， $j = 1, 2, \dots, n \times n$ 。那么对于任意一个 patch $P_j u \in \mathbb{R}^{s \times s}$ 可以对应于 $(s \times s)$ 维空间中的一个点。一张图片组成了 $(s \times s)$ 维空间中的一个点云 $\mathcal{P}(u)$ /流形 $\mathcal{M}(u)$ 。从这个观点出发，试着讨论以下问题中的一些：

1. 如果将图像处理的目标函数修改为

$$\min_u \|Au - f\|_2^2 + \lambda R(\mathcal{M}(u))$$

那么一些 patch-based 的模型(可不限于课上所讲)可以转化为什么形式的 R ?

2. 目前有些工作是基于极小化 \mathcal{M} 的维度[1,2]。能否进行复现及改进。
3. 不限于1和2，能否提出新的基于 patch 的图像处理模型?

[1] Osher, Stanley, Zuoqiang Shi, and Wei Zhu. "Low dimensional manifold model for image processing." *SIAM Journal on Imaging Sciences* 10.4 (2017): 1669-1690.

[2] Lai, Rongjie, and Jia Li. "Manifold based low-rank regularization for image restoration and semi-supervised learning." *Journal of Scientific Computing* 74.3 (2018): 1241-1263.

2 深度学习中的解映射模型

虽然深度学习在图像任务中取得了卓越的成果，但与传统方法中单图去噪不同，往往需要大量的数据集和标签 $\{u_i, f_i\}$ ，称为监督学习，对应于求解: $\min_{\theta} \sum_i \mathcal{L}(\mathcal{F}_\theta(f_i), u_i) + \hat{\mathcal{R}}(\mathcal{F}_\theta)$ 。

在实际任务中，标签数据往往难以获得。如何仅依赖数据集 $\{u_i\}$ 甚至单图 u 进行图像处理，就显得值得探究且富有挑战。无监督学习方法理论上可以写为: $\min_{\theta} \sum_i \mathcal{L}(A\mathcal{F}_\theta(f_i), f_i) + \hat{\mathcal{R}}(\mathcal{F}_\theta)$ 。

目前许多无监督的去噪算法可以写成 $\min_{\theta} \sum_i \mathcal{L}(\mathcal{F}_\theta(\hat{f}_i), \tilde{f}_i) + \hat{\mathcal{R}}(\mathcal{F}_\theta)$ 。其中 \hat{f}_i 和 \tilde{f}_i 对应于 f_i 的某种腐蚀(corruption)，目的是 $\mathcal{L}(\mathcal{F}_\theta(\hat{f}_i), \tilde{f}_i)$ 对应于 $\mathcal{L}(\mathcal{F}_\theta(f_i), u_i)$ 的无偏估计，如Noise2void [3]，Noise2Self [4]，Self2Self [5]。试着讨论以下问题中的某个：

1. 能否提出新的模型，例如引入对比学习的机制(已有工作用于去噪和去马赛克[6])。
2. 对于 A 对应于不是恒等变换的情况，如去模糊，CT图像。目前直接使用 $\min_{\theta} \sum_i \mathcal{L}(A\mathcal{F}_\theta(f_i), f_i) + \hat{\mathcal{R}}(\mathcal{F}_\theta)$ 求解的工作不多。难点在哪？是否能提出改进算法？

参考文献：

[3] Krull, Alexander, Tim-Oliver Buchholz, and Florian Jug. "Noise2void-learning denoising from single noisy images." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.

[4] Batson, Joshua, and Loic Royer. "Noise2self: Blind denoising by self-supervision." *International Conference on Machine Learning*. PMLR, 2019.

[5] Quan, Yuhui, et al. "Self2self with dropout: Learning self-supervised denoising from single image." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

[6] Dong, Nanqing, et al. "Residual Contrastive Learning for Joint Demosaicking and Denoising." *arXiv preprint arXiv:2106.10070* (2021).

3 深度学习中的图像表示模型

传统图像中的优化目标函数一般分为保真项 $\|Au - f\|$ 和正则项 $R(u)$ 。其中正则化的选取至关重要，如TV，小波系数的稀疏性等。而不同于传统图像问题，深度学习方法的正则化项往往加在网络参数，网络结构之上，或者依赖于优化算法等提供的隐式正则项。

许多方法使用 $\mathcal{U}_\theta(z)$ 来表示图像， $\|\mathcal{U}_\theta(z) - f\|_2$ 作为保真项。其中 z 为固定的随机初始化， \mathcal{U}_θ 对应于某种网络，参数为 θ 。不同方法的主要区别在于网络结构，算法的选取不同。DIP[7]使用了Early stopping的方法，背后的原理是：在 θ 的优化过程中， \mathcal{U}_θ 会先收敛至自然图像，而噪声会被抑制。Deep Decoder[8]使用了参数量很少的网络，保证 \mathcal{U}_θ 的值域接近于自然图像。BNN[9]使用了 θ 的先验分布作为正则项。试着讨论以下问题中的某个：

1. 能否分析其中的一个或几个，并提出自己的想法和改进。(例如 z 的选取是否有更加合理的方式)
2. 是否能引入新的正则化机制。如知识蒸馏(Knowledge Distillation)。动机是使用网络的输出或中间量提供给另一个网络或者自身，作为可学习目标的一部分。已有的工作将知识蒸馏的作用理解为标签光滑化[10]，early stopping[11]，正则化项[12]等。(已有工作用于图像超分辨[13,14])

参考文献：

[7] Ulyanov, Dmitry, Andrea Vedaldi, and Victor Lempitsky. "Deep image prior." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[8] Heckel, Reinhard, and Paul Hand. "Deep decoder: Concise image representations from untrained non-convolutional networks." *arXiv preprint arXiv:1810.03982* (2018). cite:116

[9] Pang, Tongyao, Yuhui Quan, and Hui Ji. "Self-supervised Bayesian Deep Learning for Image Recovery with Applications to Compressive Sensing." *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*. Springer International Publishing, 2020.

[10] Müller, Rafael, Simon Kornblith, and Geoffrey Hinton. "When does label smoothing help?." *arXiv preprint arXiv:1906.02629* (2019).

[11] Dong, Bin, et al. "Distillation \approx Early Stopping? Harvesting Dark Knowledge Utilizing Anisotropic Information Retrieval For Overparameterized Neural Network." *arXiv preprint arXiv:1910.01255* (2019)

[12] Mobahi, Hossein, Mehrdad Farajtabar, and Peter L. Bartlett. "Self-distillation amplifies regularization in hilbert space." *arXiv preprint arXiv:2002.05715* (2020).

[13] Gao, Qinquan, et al. "Image super-resolution using knowledge distillation." *Asian Conference on Computer Vision*. Springer, Cham, 2018.

[14] He, Zibin, et al. "Fakd: Feature-affinity based knowledge distillation for efficient image super-resolution." *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020.