Image Denoising survey

Monday, November 4, 2024 8:59 AM

paper: Image Denoising: The Deep Learning Revolution and Beyond — A Survey Paper by Michael Elad, Bahjat Kawar, Gregory Vaksman
SIAM J. Imaging Science, 2023

① What is image denoising?

$$y = x + v, \quad \text{all vectors in } \mathbb{R}^N \quad (\text{eg. } N = n_x \times n_y)$$

y observed, x true image, v noise

* note we wlog use vectors not matrices

It's the simplest inverse problem ($y = A \cdot x + v$) ie. $A = I$

- Assume $v \sim N(0, \sigma^2 I)$ and σ^2 known. } "AWGN" Additive, white Gaussian Noise

  ... the simplest noise model.

  → real cameras (CCD pixel arrays) have a tiny amount of AWGN. Mostly "shot noise" due to quantum nature of discrete photons, following a Poisson distribution
  (partly addressed via Anscombe or other variance stabilizing distribution, model as Gaussian if photon count high)

- Why so simple?
  - Fundamental: backbone of more realistic setups ($A \neq I$, and other noise) } indirect
  - Tractable, understand
  - Paper argues its a key ingredient in fancier setups } direct

- Goal

  Build a denoiser D, return estimate

$$\hat{x} = D(y, \sigma)$$

Image Denoising survey (p. 2)

Monday, November 4, 2024 9:11 AM

- What criterion?

For now, for good reason, use MSE (equivalently, PSNR)

*Careful: before, "MSE" is $\underbrace{\frac{1}{\#\text{pixels}}}_{\text{actual constant of little importance if it never changes}} \underbrace{\sum_{i=1}^{\#\text{pixels}} (\hat{x}_i - x)^2}_{\|\hat{x}-x\|_2^2}$

and "mean" referred to this average

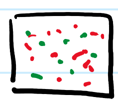
Now,

$$MSE = \mathbb{E}\,\|\hat{x}-x\|_2^2 := \int \|\hat{x}-x\|_2^2 \cdot p(x)\,dx$$

$(\hat{x} = D(y,\sigma),\ y = x+v,$ so $\hat{x}$ is a function of $x)$

i.e. "mean" is used in a different sense.

That is, we're assigning a prior distribution to the set of images.

$p(X =$  $) = 0.013$, $p(X =$  $) = 10^{-20}$

This is the critical ingredient...

without a prior, there's no hope

Setting $\hat{x} = y$ is the best you can do.

So this is Bayesian? not Frequentist?

Note to future years: lecture was 1 day before 2024 presidential election

Yes, sort of...

but we don't (anymore) write down a formula for $p(X)$

(these days, we'll learn it from our dataset)

Image Denoising survey (p. 3)

Monday, November 4, 2024 9:21 AM

Estimation 101

$y = x + v$, P_X is prior on X

$P_{X,Y}$ is joint on $X \times Y$

whether Bayesian or not, Bayes' rule is true:

$$P_{X|Y=y}(x) = \frac{P_{X,Y}(x,y)}{P_Y(y)} = \frac{P_{Y|X=x}(y)\, P_X(x)}{P_Y(y)}$$

"X" refers to r.v.
"x" refers to possible value of r.v.

Maximum Likelihood Estimation (MLE) [not "Bayesian"] very popular

Find x to maximize $p(\underline{y}|x)$ ("likelihood") or minimize negative log likelihood

Given x, $y = x + v$ i.e. $y \sim N(x, \sigma^2 I)$

so $p(y|x) = \text{const} \cdot \exp(-\|y-x\|^2/2\sigma^2)$

So

$$\hat{x}_{MLE} = \underset{x}{\operatorname{argmin}} -\log(p(y|x)) = \underset{x}{\operatorname{argmin}} \frac{1}{2}\frac{1}{\sigma^2}\|y-x\|^2 = y$$

$\hat{x}_{MLE} = y$ not useful! (if $A \neq I$ it can be useful)

Maximum 'a posteriori' Estimation (MAP) [Bayesian]

Find x to maximize $p(x|\underline{y})$

$$p(x|y) = \frac{p(y|x)\,p(x)}{\cancel{p(y)}}$$ unimportant

$$\min_x -\log(p(x|y)) = \min_x \underbrace{\underbrace{-\log(p(y|x))}_{\text{MLE term}}}_{\text{data fitting term}} \underbrace{\overbrace{\underbrace{-\log(p(x))}_{\rho(x)}}^{\text{if } p(x) = \text{constant (i.e. uniform, uninformative)}}}_{\text{regularization term}}$$

then this has no effect so MLE = MAP in this case

exploits a prior

assume $p(x) \sim e^{-\rho(x)}$

Image Denoising survey (p. 4)

Monday, November 4, 2024 11:02 AM

Ex suppose $p(x) = \text{const} \cdot \exp(\overbrace{-\|x\|_2^2}^{\rho(x)})$ i.e. we prioritize small values of pixels

then MAP:

$$\min_x -\log(p(y|x)) - \log(p(x))$$

$$= \min_x \frac{1}{2\sigma^2}\|x-y\|^2 + \frac{1}{2}\|x\|^2$$

Tikhonov regularization / Ridge regression

MMSE minimum MSE estimator [Bayesian]

Find $\hat{x}$ to minimize MSE $\mathbb{E}\|\hat{x}-x\|^2$

$$\hat{x}_{MMSE} = \mathbb{E}[x|y]$$

"Closed form" misleading

not helpful directly except in simplest cases

(2) "Classical" Techniques

... most (not all) involved choosing priors $p(x)$, or, equivalently, choosing $-\log(p(x)) =: \rho(x)$