Noise2Noise

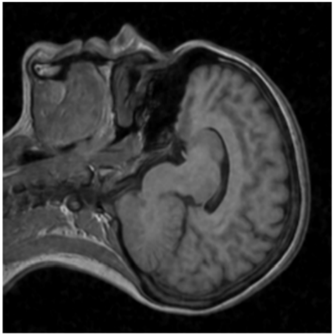
Хрушков Павел

НИУ ВШЭ

26 сентября 2018 г.

Denoising

Задача: дано зашумленное изображение, нужно выдать «чистое» изображение.
Проблема: для обучения нужны пары (зашумленное изображение, «чистое» изображение)



Важное наблюдение

Пусть x — объект, а y — целевая переменная (случайная величина), и мы строим модель $f(x)$. Тогда, как мы знаем,

$$\arg\min_{f(x)} \mathbb{E}_{(x,y)} [f(x) - y]^2 = \mathbb{E}(y|x)$$

Пусть теперь $\hat{y} = y + \varepsilon$, ε — шум (независимый от x), для которого $\mathbb{E}\varepsilon = 0$. Тогда

$$\arg\min_{f(x)} \mathbb{E}_{(x,y)} [f(x) - y]^2 = \arg\min_{f(x)} \mathbb{E}_{(x,\hat{y})} [f(x) - \hat{y}]^2$$

Важное наблюдение

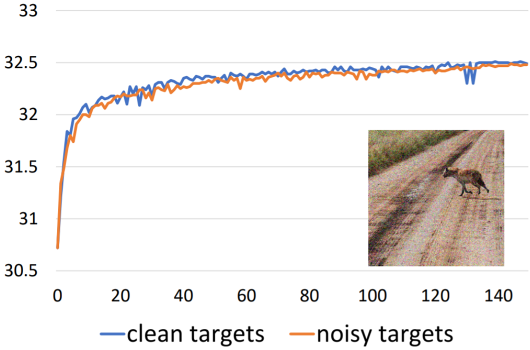
Вывод

Зашумление таргета шумом с нулевым матожиданием не портит модель (с точки зрения квадратичного эмпирического риска)

Не стоит думать, что рассуждения справедливы только для квадратичной функции потерь. Например, если мы при оптимизации *абсолютной* функции потерь заменим y на $\hat{y}$ такие, что $\mathrm{median}(y|x) = \mathrm{median}(\hat{y}|x)$, мы тоже не испортим модель.

Эксперимент 1: гауссовский шум

Возьмем 50'000 картинок из ImageNet, зашумим их. В качестве таргета для очередного изображения при обучении будем давать то же изображение, но зашумленное по-другому.



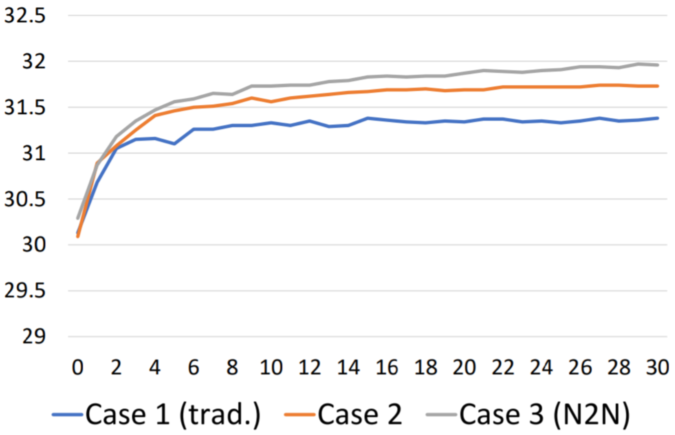
Реальность: фиксированный бюджет

В реальности у нас есть только доступ к по-разному зашумленным версиям одного и того же изображения. Обозначим

M — число по-разному зашумленных версий одного изображения
N — число различных изображений
$NM = 2000$ — фиксированное число картинок

1. $N = 100$, $M = 20$, фиксированный объект, 19 таргетов
2. $N = 100$, $M = 20$, пары объект-таргет — все возможные сочетания зашумленных картинок по две
3. $N = 1000$, $M = 2$

Эксперимент 2: фиксированный бюджет



Результаты: гауссовский шум

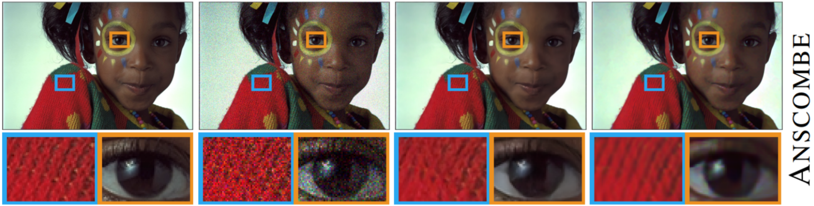
Пуассоновский шум

Частый шум на фотографиях — пуассоновский. Моделируется он следующим образом.

1. Берем изображение. MinMaxScale'им пиксели так, чтобы они попали в интервал $[0; \lambda]$.
2. Для каждого пикселя x с интенсивностью $\lambda(x)$ сэмплируем новое зашумленное значение $\sim \text{Pois}(\lambda(x))$.
3. Производим обратное MinMaxScale преобразование.

Так как матожидание пуассоновской величины с параметром λ равно λ, получим шум с нулевым средним.

Результаты: пуассоновский шум



Биномиальный шум

Рассмотрим случайную величину $m \sim \mathrm{Ber}(p)$. Пусть $\hat{y} = my$, тогда

$$\begin{aligned}
\arg\min_{f(x)} \mathbb{E}_{(m,x,\hat{y})} m[f(x) - \hat{y}]^2 &= \arg\min_{f(x)} \mathbb{E}_{(m,x,y)} m[f(x) - y]^2 \\
&= \arg\min_{f(x)} p\, \mathbb{E}_{(x,y)} [f(x) - y]^2 \\
&= \arg\min_{f(x)} \mathbb{E}_{(x,y)} [f(x) - y]^2
\end{aligned}$$

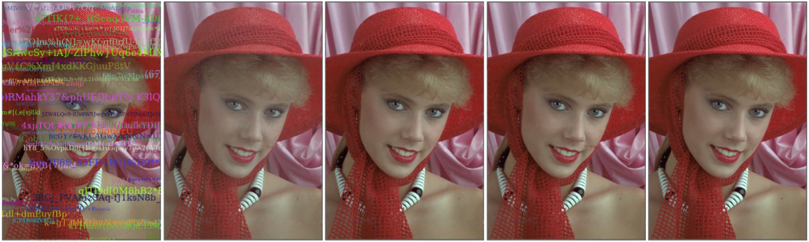
Итак, если мы будем вносить шум в y с помощью случайной бернуллиевой маски m и оптимизировать функцию потерь $m[f(x) - y]^2$, мы опять не испортим модель.

Результаты: биномиальный шум

Сравнение методов при различных шумах

	Gaussian (σ=25)			Poisson (λ=30)			Bernoulli (p=0.5)		
	clean	noisy	BM3D	clean	noisy	ANSC	clean	noisy	DIP
Kodak	32.50	32.48	31.82	31.52	31.50	29.15	33.01	33.17	30.78
BSD300	31.07	31.06	30.34	30.18	30.16	27.56	31.04	31.16	28.97
Set14	31.31	31.28	30.50	30.07	30.06	28.36	31.51	31.72	30.67
Average	**31.63**	**31.61**	**30.89**	**30.59**	**30.57**	**28.36**	**31.85**	**32.02**	**30.14**

Будем зашумлять изображение, добавляя на него текст.
Предположение: медиана пикселя совпадает с реальным значением пикселя.



Input ($p \approx 0.25$)	L_2	L_1	Clean targets	Ground truth
17.12 dB	26.89 dB	35.75 dB	35.82 dB	PSNR

Impulse noise

Рассмотрим расширение биномиального шума. Опять реализуем $m \sim \mathrm{Ber}(p)$. Если $m = 1$, положим $\hat{y} = y$, иначе $\hat{y}$ присвоим случайное значение (из равномерного распределения). Для $0 \le t \le 1$

$$F_{\hat{y}}(t) = pF_y(t) + (1-p)t \Rightarrow \varrho_{\hat{y}}(t) = p\varrho_y(t) + (1-p)$$

А это значит, что $\arg\max_t \varrho_{\hat{y}}(t) = \arg\max_t \varrho_y(t)$, т.е. у y и $\hat{y}$ совпадают моды.

Нам нужна такая функция потерь, оптимальной константной для которой являлась бы мода распределения.

ℓ_0-loss

Пусть $z(p)$ — оптимальная константа для задачи минимизации, соответствующей ℓ_p-норме:

$$z(p) = \arg\min_z \mathbb{E}_y \left[\frac{1}{p}|z - y|^p\right]$$

Оказывается, существует предел $\lim_{p\to 0} z(p)$, который обычно близок к моде распределения y. Авторы статьи называют поиск $z(0)$ оптимизацией ℓ_0-loss.

На практике оптимизацию ℓ_0-loss аппроксимируют, оптимизируя $\hat{L}_\gamma(z, y) = (|z - y| + \varepsilon)^\gamma$, где γ линейно уменьшается от 2 до 0 по мере обучения.

Интуиция: ℓ_0-«норма»

Пусть y дискретная случайная величина. Тогда

$$\begin{aligned}\mathrm{mode}(y|x) &= \arg\min_{f(x)} \mathbb{E}_x[1 - \mathbb{P}(y = f(x)|x)] \\ &= \arg\min_{f(x)} \mathbb{E}_x \mathbb{E}_y([y \neq f(x)]|x) \\ &= \arg\min_{f(x)} \mathbb{E}_{(x,y)}[y \neq f(x)]\end{aligned}$$

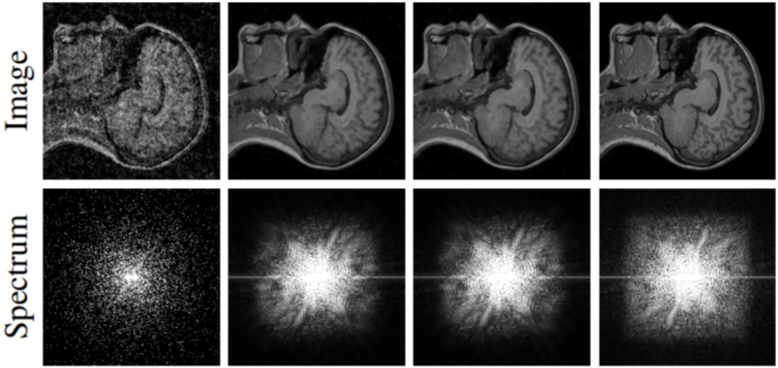
Определение

ℓ_0-«нормой» (в кавычках!) вектора $x \in \mathbb{R}^n$ называют число его ненулевых координат: $\sum_{i=1}^n [x_i \neq 0] = \sum_{i=1}^n |x_i|^0$. $(0^0 := 0)$

Результаты: ℓ_0-«loss»

Input ($p = 0.70$)	L_2 / L_1	L_0	Clean targets	Ground truth
8.89 dB	13.02 dB / 16.36 dB	28.43 dB	28.86 dB	PSNR

И так далее...



(a) Input 18.93 dB (b) Noisy trg. 29.77 dB (c) Clean trg. 29.81 dB (d) Reference

Статья

Noise2Noise: Learning Image Restoration without Clean Data