

Байесовские SRGAN

Егор Крашенинников

Физико-техническая школа СПбАУ НОЦНТ РАН

Научный руководитель: Александр Сироткин

E-mail: krashennikovvegor@gmail.com.

Аннотация

В этой работе исследуется применение Байесовской интерпретации генеративных состязательных сетей к задаче увеличения разрешения изображений.

1 Введение

1.1 Классические GAN

Генеративные состязательные сети (GANs), впервые представленные Яном Гудфеллоу в 2014 году [1], находят применение в задачах генерации искусственных данных. GAN состоит из двух моделей – генератора G и дискриминатора D . Генератор пытается создавать примеры, как можно более похожие на реальные, а дискриминатор – как можно лучше отличать реальные примеры от сгенерированных. Несмотря на всю эффективность GAN-ов, их обучение часто ведет к эффекту коллапсирования мод, когда генератор просто запоминает несколько примеров, чтобы обмануть дискриминатор.

1.2 Байесовские GAN

Насколько мы знаем, Юнус Саатчи и Эндрю Гордон Вильсон первыми представили детализированную Байесовскую интерпретацию GAN [2]. Они утверждают, что она способна избежать проблемы коллапсирования мод, а также, что методы Байеса позволяют добиться более высокой генерализации модели и превзойти классическую версию. В своей работе они вводят априорные распределения на параметры генератора и дискриминатора, а маргинализацию апостериорных распределений над этими параметрами совершают, используя стохастические градиенты НМС (Hamiltonian Monte Carlo) [3].

На рисунке 1, взятом из [2], изображено мультимодальное апостериорное распределение параметров генератора, каждая конфигурация которых соответствует собственной генеративной гипотезе. Можно увидеть два примера изображений рукописных цифр,

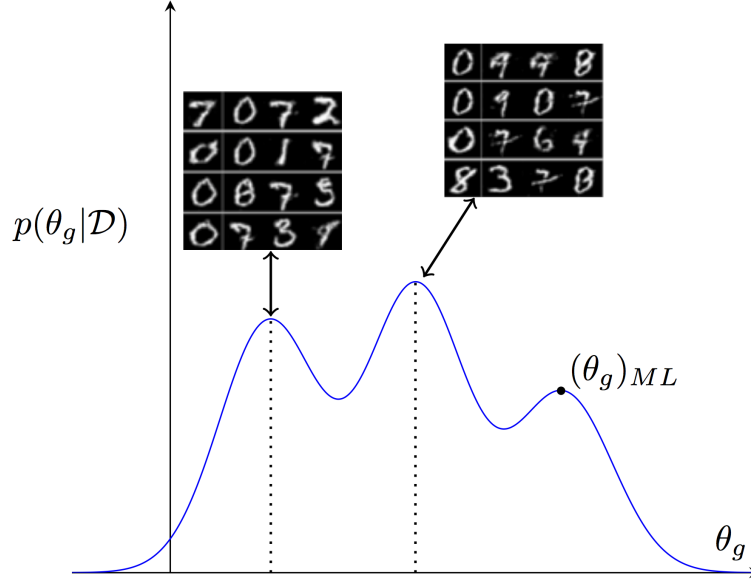


Рис. 1

сгенерированных исходя из двух различных конфигураций. Также наглядно показано, что, в отличие от Байесовских GAN, классические производят лишь точечную оценку всего распределения параметров методом максимального правдоподобия (ML) и тем самым теряют потенциально значимые интерпретации данных.

2 Байесовские SRGAN

Одной из задач, в которой классические GAN-ы добились больших успехов, является задача увеличения разрешения изображений [4]. Мы предлагаем использовать недавно представленные Байесовские GAN для большей натурализации генерируемых изображений повышенного разрешения. Иными словами, наша цель – сделать модель генератора более естественной.

2.1 Метод

Имея набор изображений \mathcal{D} , допустим, что I^{LR} – зашумленное нормальным шумом изображение пониженного разрешения, полученное в результате применения перечисленных операций к соответствующей I^{HR} , $I^{HR} \in \mathcal{D}$. Введем генератор и дискриминатор, как $G(I^{LR}; \theta_g)$ и $D(C; \theta_d)$, параметризованные θ_g и θ_d соответственно; $\theta_g \sim p(\theta_g)$, $\theta_d \sim p(\theta_d)$, C принимает значения либо I^{HR} , либо $G(I^{LR}; \theta_g)$, в зависимости от тренировочного шага. Пусть L – некая функция схожести между двумя изображениями. Она необходима для того, чтобы генератор учился создавать изображения, соответствующие его входным.

Объявив распределения над распределениями параметров, мы получаем континуум всех возможных генераторов и дискриминаторов. Апостериорный вывод выглядит следующим образом:

$$p(\theta_g|X, \theta_d) \propto \prod_{i=1}^n D(G(I_i^{LR}; \theta_g); \theta_d) \times \prod_{i=1}^n L(G(I_i^{LR}; \theta_g); I_i^{HR}) \times p(\theta_g|\alpha_g)$$

$$p(\theta_d|X, \theta_g) \propto \prod_{i=1}^n D(I_i^{HR}; \theta_d) \times \prod_{i=1}^n (1 - D(G(I_i^{LR}; \theta_g); \theta_d)) \times p(\theta_d|\alpha_d)$$

где $X = \{I_i^{LR}, I_i^{HR}\}_{i=1}^n$ – набор из n тренировочных примеров, а α_g и α_d – соответствующие гиперпараметры априорных распределений.

Следуя Bayesian GAN [2] маргинализация апостериорных распределений выглядит так:

$$p(\theta_g|\theta_d) = \int p(\theta_g, X|\theta_d) dX = \int p(\theta_g|X, \theta_d) \overbrace{p(X|\theta_d)}^{=p(X)} dX \approx \frac{1}{J_g} \sum_{i=1}^{J_g} p(\theta_g|X_i, \theta_d)$$

Аналогично,

$$p(\theta_d|\theta_g) \approx \frac{1}{J_d} \sum_{j=1}^{J_d} p(\theta_d|X_j, \theta_g)$$

Таким образом, мы сэмплируем примеры из распределения изображений, обуславливающие апостериорный вывод распределений $p(\theta_g|X, \theta_d)$ и $p(\theta_d|X, \theta_g)$, с помощью которых аппроксимируются $p(\theta_g|\theta_d)$ и $p(\theta_d|\theta_g)$.

3 Обучение

Тренировочные примеры изображений были взяты из данных ImageNet [5]. В отличие от архитектуры Bayesian GAN [2], BSRGAN состоит из более глубоких моделей генератора и дискриминатора. Это снижает скорость обучения и увеличивает объем требуемой памяти, но при этом, для большей скорости сходимости, в генераторе присутствуют *residual blocks* [6]. Для того, чтобы генерируемые изображения соответствовали входным изображениям пониженного разрешения, мы считаем разницу выделенных моделью VGG19 [7] признаков из оригинальных картинок и созданных генератором, а затем используем эту разницу в качестве добавки к функции ошибки. На рисунке 2 показано восстановление одного из примеров после лишь десяти тысяч итераций обучения модели. Слева – зашумленное изображение пониженного разрешения, по центру – сгенерированное одним из генераторов (64x64), справа – исходное.

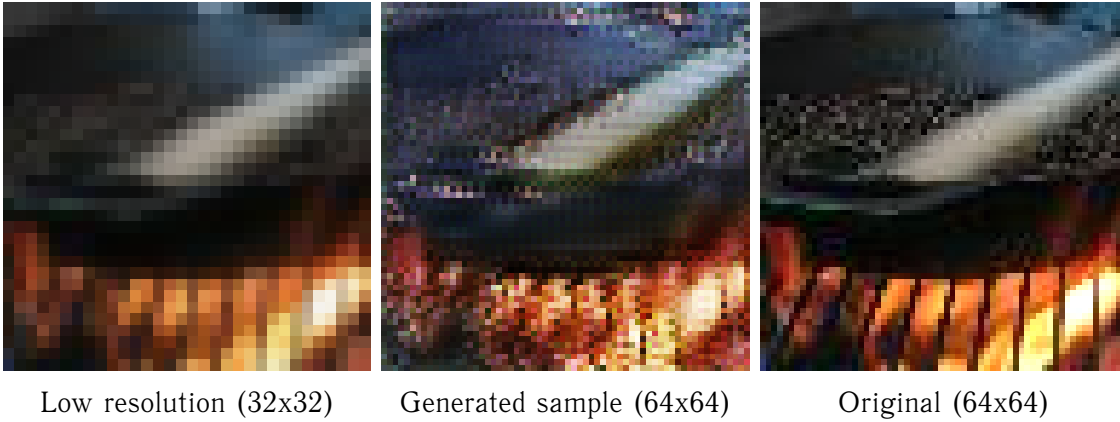


Рис. 2

4 Результат

Было показано, что предложенный метод **потенциально** способен дать хорошие результаты в задаче увеличения разрешения изображений. Таким образом, сходимость наблюдается, но подобрать правильные гиперпараметры – это еще одна достаточно сложная задача, которую мы в дальнейшем будем пытаться решить.

Список литературы

- [1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair et al., *Generative Adversarial Networks*, *ArXiv e-prints* (2014) [1406.2661].
- [2] Y. Saatchi and A. G. Wilson, *Bayesian GAN*, *ArXiv e-prints* (2017) [1705.09558].
- [3] T. Chen, E. B. Fox and C. Guestrin, *Stochastic Gradient Hamiltonian Monte Carlo*, *ArXiv e-prints* (2014) [1402.4102].
- [4] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta et al., *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*, *ArXiv e-prints* (2016) [1609.04802].
- [5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, *ImageNet: A Large-Scale Hierarchical Image Database*, in *CVPR09*, 2009.
- [6] K. He, X. Zhang, S. Ren and J. Sun, *Deep Residual Learning for Image Recognition*, *ArXiv e-prints* (2015) [1512.03385].
- [7] K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, *ArXiv e-prints* (2014) [1409.1556].