

Исследование обобщающей способности методов глубокого обучения для улучшения качества изображений

Н. И. Бережнов, email: beregnovnikita@yandex.ru

А. А. Сирота, email: sir@cs.vsu.ru

Воронежский государственный университет

Аннотация. В работе исследуются методы глубокого обучения в области улучшения качества изображения. Под улучшением качества изображений здесь понимается: восстановление изображений при наличии дефектов различной природы, увеличение резкости изображений, устранение эффектов съемки в условиях плохих погодных условий. В работе описываются современные архитектуры глубоких нейронных сетей, такие как сверточные нейронные сети, модели-трансформеры, архитектуры с механизмом внимания. Используется три датасета. Датасет ImageNet (50000 изображений), с наложенными на него разными искажениями в виде аппликативных и аддитивных помех и размытия. SIDD – датасет уже зашумлённых изображений с имеющимися эталонами. Погодный датасет с изображениями снега, дождя, тумана, где также имеются эталонные изображения. Для сравнения качества восстановления изображений используются следующие метрики: *PNSR*, *SSIM*, *FID*. Показано, что обобщающая способность нейронных сетей сильно зависит от исходного датасета, использованного при обучении. Обученные модели показывают существенно худшие результаты на незнакомых им типах изображений. Эти результаты можно улучшить путем дообучения нейронной сети конкретно под определённую задачу. Проблема создания универсального алгоритма для повышения качества изображений остаётся открытой.

Ключевые слова: Повышение качества изображений, нейронные сети, глубокое обучение, модели-трансформеры, восстановление изображений.

Введение

На сегодняшний день существует много работ в области улучшения качества изображений. В последнее время становится популярным использование архитектур с модулями внимания, генеративно-состязательных сетей (GAN), а также моделей-трансформеров и диффузных моделей. Последние набирают всё больше популярность. В

частности, они даже смогли обойти GAN в задаче построения сверхразрешения. В большинстве случаев исследователи сосредотачиваются на конкретной практической задаче, оптимизируя параметры нейронной сети под неё. Однако после этого нейронная сеть теряет свою обобщающую способность. В тоже время встречаются работы [1,2], где авторы проводят исследование архитектуры нейронной сети на многозадачность. В работе [1] авторы используют диффузную модель, которая через итеративный процесс обучается восстанавливать и экстраполировать до панорамного вида изображения различного типа. При этом используются мощные GPU для обучения модели, а также большие батчи данных.

Как известно, глубокое обучение требует больших наборов данных. К сожалению, сложно собрать большой датасет испорченных данных вместе с истинными чистыми изображениями. Исследователи [2] предлагают напрямую использовать зашумлённые изображения в качестве входа и выхода нейронной сети. Утверждается, что статистически они будут стремиться к истинным изображениям. Однако данный подход проигрывает в точности нейронным сетям с большим числом параметров. В том числе и новому тренду в компьютерном зрении – vision transformer (ViT). В виду ограниченности данных в работе применялась генерация дефектов на изображениях, а также использовались готовые датасеты с уже имеющимися зашумлёнными изображениями и эталонами [3][4]. В данной работе предлагается рассмотреть следующие искажения: размытие изображения, искажение импульсными, аппликативными и аддитивными помехами, влияние погодных осадков. Последние были выбраны из-за того, что данный вид искажений имеет сложную природу и является реальным применением задачи восстановления изображений на практике.

В связи с этим в данной работе решено было выбрать несколько архитектур глубоких нейронных сетей и обучить их на датасетах различной природы. Целью данной работы явилось – проверить данные архитектуры глубоких нейронных сетей на предмет их универсальности и обобщающей способности.

1. Сверточные нейронные сети

С увеличением мощностей графических процессоров стало возможно обучать глубокие нейронные сети с большим количеством слоёв. Сверточные сети в начале 2010-х годов совершили революцию в компьютерном зрении. Во-первых, с помощью свёрток удалось значительно уменьшить количество параметров, нежели чем использовать MLP. Во-вторых, они обладают свойством пространственной инвариантности – независимо от того, где

расположен объект на изображении, свертка сможет его найти. Также исследователям удалось решить проблему затухающего градиента путём добавления остаточных связей или так называемых skip-connections.

В связи с этим появилось множество архитектур, содержащих до нескольких сотен сверточных слоёв. Эти же изменения пришли и в область восстановления изображений. Исследователи [5] используют глубокую нейронную сеть REDNet с остаточными связями. Данная архитектура обладает простотой настройки и показывает неплохие результаты в области улучшения качества изображений. В работе решено было использовать именно её с 15 сверточными слоями.

2. Модели-трансформеры

Модели-трансформеры впервые появились в задачах, связанных с обработкой естественного языка. Однако в связи с успехом языковых моделей Bert и GPT-3 исследователи задумались над использованием их в компьютерном зрении. Первым шагом в эту сторону стала модель vision transformer (ViT). Она состоит из трёх частей: линейных проекционных уровней, кодировщика и головы. Изображение разбивается на n одинаковых частей, затем каждая из них проходит через линейные проекционные уровни. Далее выполняется позиционное кодирование для того, чтобы понимать взаимное расположение частей изображения. После этого добавляется дополнительный обучаемый патч, необходимый для извлечения признаков из изображения. Все патчи проходят через кодировщик, состоящий из полносвязных слоёв и механизма внимания. На выходе из которого будет $n + 1$ частей, однако, самая важная из них будет первая, она называется головой. В дальнейшем её можно использовать для любых задач машинного обучения. На рис. 1 представлена архитектура ViT для задачи классификации изображений.

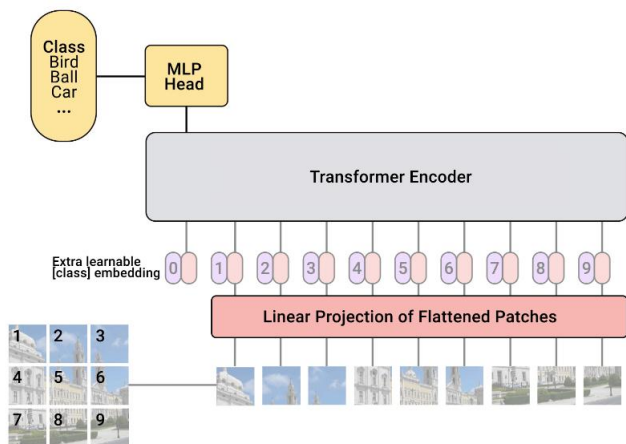


Рис. 1. Архитектура vision transformer

В работе будет использоваться архитектура SUNet [6], построенная на подобию и принципах архитектуры UNet. Авторы оставляют только сверточные слои на входе и на выходе, а также блоки повышения и понижения дискретизации, все остальные блоки – трансформеры.

3. Механизмы внимания

Рассмотрим подробнее существующие механизмы внимания. Основной принцип внимания таков: имеются ключи K , значения V и запросы Q . Задача сводится к тому, чтобы по запросам Q определить релевантные значения V . Мерой близости служит скалярное произведение между Q и K . Схема механизма внимания представлена на рис. 2.

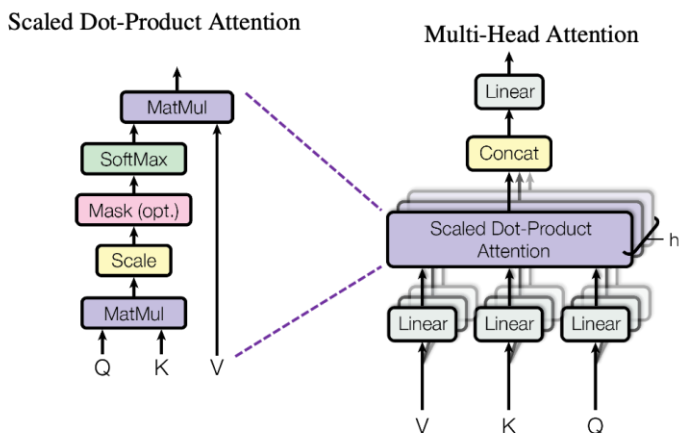


Рис. 2. Механизм внимания

Представленный выше механизм внимания обладает большой вычислительной сложностью, поэтому существует большое число его модификаций. Очень часто используется каналные, пространственные и локальные механизмы внимания.

Механизмы внимания оказались настолько эффективны в трансформерах, что стали активно использоваться в сверточных нейронных сетях. Например, в архитектуре MIRNet [7] авторы используют каналное и пространственное внимание, исследуют принцип dual attention. В работе решено было выбрать данную нейронную сеть в реализации от keras.io [8].

4. Метрики для сравнения качества изображений

Для сравнения качества двух изображений: эталона и улучшенного изображения использовались следующие метрики: SSIM (чем больше, тем лучше) PNSR (чем больше, тем лучше) и FID (чем меньше, тем лучше). SSIM (коэффициент структурного сходства изображений) и PNSR (пиковое отношение сигнал-шум) являются классическими метриками в области обработки изображений. Однако многими исследователями было показано, что они не всегда соответствуют человеческому восприятию изображения. Поэтому решено было использовать ещё метрику Fréchet Inception Distance (FID). Она основана на расстоянии Фреше между двумя распределениями признаков улучшенных и эталонных изображений. По сути, эта метрика показывает насколько одно многомерное распределение похоже на

другое. Для выделения признаков из изображений используется нейронная сеть InceptionV3, обученная на датасете ImageNet. В явном виде формулу можно записать следующим образом:

$$FID = \sum (\mu_1 - \mu_2)^2 + Tr(C_1 + C_2 - 2 \cdot \sqrt{C_1 C_2}) \quad (1)$$

где μ_1 , μ_2 - математические ожидания соответствующих распределений, сумма берётся по всем пикселям, а C_1 , C_2 - матрицы ковариаций, составленные между всеми элементами данных распределений.

Из формулы можно увидеть: чем меньше значение, тем ближе друг к другу распределения. Данная метрика уже зарекомендовала себя в различных задачах компьютерного зрения, поэтому будем оценивать ею распределения, полученные из тестовой выборки, состоящей примерно из 500 изображений для каждого датасета.

5. Обучение и результаты

В работе для исследования задачи улучшения изображений используются три датасета. Датасет ImageNet (50000 изображений), с наложенными на него различными видами искажений в виде аппликативных, аддитивных помех и размытия. Данные искажения подробно описаны в работе авторов [9]. Помимо этого, использовался SIDD датасет с уже зашумлёнными изображениями, снятыми на смартфон, и имеющимися эталонными изображениями, а также датасет изображений, снятых в плохих погодных условиях (снег, дождь, туман), где также имеются эталонные изображения. Обучение происходило на видеокарте NVIDIA GeForce RTX 2070 с 8 гигабайтами видеопамяти.

В работе решено было использовать описанные выше архитектуры REDNet, MIRNet и SUnet следующим образом:

- обучение и тестирование на одном датасете;
- обучение на одном датасете и тестирование на другом;
- обучение на одном датасете, затем перенос обучения и тестирование на другом датасете.

Результаты оценки выбранных метрик на тестовой выборке приведены в таблицах ниже. Лучшие значения показателей выделены полужирным шрифтом.

Таблица 1

Обучение и тестирование на одном датасете

Нейронная сеть	ImageNet			SIDD			Weather dataset		
	PSNR	SSIM	FID	PSNR	SSIM	FID	PSNR	SSIM	FID
REDNet	18.91	0.69	104.0	32.02	0.84	81.10	20.70	0.80	134.1
MIRNet	22.92	0.71	106.0	32.06	0.84	78.29	20.98	0.81	142.8
SUNet	25.98	0.79	55.60	39.79	0.96	61.32	20.21	0.69	92.42

Таблица 2

Обучение на ImageNet и тестирование на остальных

Нейронная сеть	SIDD			Weather dataset		
	PSNR	SSIM	FID	PSNR	SSIM	FID
REDNet	22.67	0.79	90.65	18.91	0.76	158.8
MIRNet	31.94	0.84	82.46	18.00	0.72	166.0
SUNet	32.99	0.80	70.48	21.16	0.76	85.05

Таблица 3

Обучение на SIDD и тестирование на остальных

Нейронная сеть	ImageNet			Weather dataset		
	PSNR	SSIM	FID	PSNR	SSIM	FID
REDNet	20.88	0.70	106.8	17.60	0.70	165.2
MIRNet	20.12	0.68	108.3	17.27	0.68	169.2
SUNet	20.38	0.58	118.3	19.96	0.61	124.8

Таблица 4

Обучение на датасете плохих погодных условий и тестирование на остальных

Нейронная сеть	ImageNet			SIDD		
	PSNR	SSIM	FID	PSNR	SSIM	FID
REDNet	16.76	0.65	115	22.14	0.77	93.87
MIRNet	16.40	0.63	109	26.38	0.80	83.71
SUNet	18.60	0.60	85.87	23.30	0.66	112.9

Результаты, представленные в табл. 4 показали, что качество изображений сильно падает при тестировании на датасете изображений, снятых в плохих погодных условиях. Это можно объяснить тем, что помехи в нём обладают более сложной природой и необходимо дополнительное обучение нейронной сети. В тоже время, можно отметить, что на датасете SIDD неплохо справляются модели, обученные на ImageNet. Это связано с тем, что одним из видов искажений на ImageNet был аддитивный гауссовский шум, который похож на искажения датасета SIDD.

Перенос обучения решено было делать только с предобученной сетью на датасете ImageNet, так как он обладает наибольшей степенью общности и сложности применяемых к нему искажений.

Таблица 5

Обучение на ImageNet и затем перенос обучения и тестирование на остальных

Нейронная сеть	SIDD			Weather dataset		
	PSNR	SSIM	FID	PSNR	SSIM	FID
REDNet	32.07	0.84	84.84	18.87	0.77	149.5
MIRNet	32.09	0.84	82.88	19.49	0.76	156.7
SUNet	38.72	0.95	45.70	21.79	0.76	85.86

Визуальные примеры результатов улучшения качества изображений представлены на рис. 3.



Рис. 3. Результаты улучшения качества изображений. Слева – истинное изображение, по центру – зашумлённое и справа – улучшенное

Заключение

Проведено сравнение архитектур нейронных сетей. Исследованы возможности их обобщающей способности и универсальности применения на различных датасетах. Показано, что качество нейронных сетей сильно падает при тестировании на незнакомом датасете. Лучшее качество было получено, когда нейронная сеть обучалась и тестировалась на одном и том же датасете. Стоит отметить, что в некоторых случаях качество восстановления изображений на других датасетах падало не очень сильно, если использовался перенос обучения. Однако вопрос об универсальности нейросетевого алгоритма остаётся открытым и послужит темой дальнейших исследований авторов.

Список литературы

1. Saharia, C. Palette: Image-to-Image Diffusion Models / C. Saharia, W. Chan, H. Chang // ACM SIGGRAPH 2022 Conference Proceedings. – 2022. – DOI: 10.1145/3528233.3530757.

2. Lehtinen, J. Noise2Noise: Learning Image Restoration without Clean Data/ J. Lehtinen, J. Munkberg, J. Hasselgren // arXiv: 1803.04189. – 2018.
3. SIDD (Smartphone Image Denoising Dataset) [Электронный ресурс]. – Режим доступа : <https://paperswithcode.com/dataset/sidd>
4. TransWeather [Электронный ресурс]. – Режим доступа : <https://github.com/jeaya-maria-jose/TransWeather>
5. Xiao-Jiao, M. Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections / M. Xiao-Jiao, S. Chunhua, Y. Yubin // NIPS. – 2016.
6. Chi-Mao, F. SUNet: Swin Transformer UNet for Image Denoising / F. Chi-Mao, L. Tsung-Jung, L. Kuan-Hsien // ISCAS. – 2022. – DOI: 10.1109/ISCAS48785.2022.9937486.
7. Zamir, S. Learning Enriched Features for Fast Image Restoration and Enhancement / S. Zamir, A. Arora, S. Khan // TPAMI. – 2020. – DOI: 10.1109/TPAMI.2022.3167175.
8. Low-light image enhancement using MIRNet [Электронный ресурс]. – Режим доступа : <https://keras.io/examples/vision/mirnet/>
9. Бережнов, Н.И. Универсальный алгоритм улучшения изображений с использованием глубоких нейронных сетей / Н.И. Бережнов, А.А. Сирота // Вестник ВГУ. Серия: Системный анализ и информационные технологии. – 2022. – №2 – С. 81-92.