

РАЗРАБОТКА АЛГОРИТМА ЧАСТИЧНОГО СЖАТИЯ ТЕКСТУР С ИСПОЛЬЗОВАНИЕМ НЕЙРОННОЙ СЕТИ SRCNN

Гунько А.А.¹, Коровкин В.А.²

¹ НИ ТПУ, г. Томск (студент гр. 8ИМ31 ОИТ, ИШИТР), e-mail: aag131@tpu.ru

² НИ ТПУ, г. Томск (старший преподаватель ОИТ, ИШИТР), email: alcasar@tpu.ru

Аннотация

В докладе рассматриваются способы объединения работы алгоритма сжатия текстур S3TC и свёрточной нейронной сети SRCNN. Было проведено тестирование нескольких моделей SRCNN на различных изображениях. После анализа результатов был сделан вывод о каждой модели и определен оптимальный способ объединения S3TC и SRCNN.

Ключевые слова: текстура, сжатие, алгоритм, игровой движок, нейросеть.

Введение

Разработка приложений виртуальной и дополненной реальности позволяет добиться высокого качества и детализации визуального контента. Однако использование реалистичных текстур с высоким разрешением для объектов игрового мира и окружения существенно увеличивает объем хранимых и передаваемых данных. Это может затруднять разработку и эксплуатацию приложений [1].

Актуальность данной работы обусловлена необходимостью расширить возможности алгоритмов сжатия текстур с помощью применения нейронных сетей. Это решение позволит обеспечить баланс между производительностью и визуальным качеством при разработке приложений на игровых движках [2].

Одним из современных подходов к восстановлению детализации изображений является метод супер-разрешения – увеличение разрешения изображения с восстановлением мелких деталей на основе обученных моделей. В данной работе проводится исследование и поиск подхода, который объединит в себе работу нейронной сети SRCNN по восстановлению качества изображения и алгоритма частичного сжатия текстур на примере S3TC [1, 2].

Целью работы является разработка алгоритма частичного сжатия текстур с восстановлением качества посредством нейронной сети SRCNN.

Для достижения поставленной цели необходимо выполнить следующие задачи:

1. исследовать метод супер-разрешения изображений с использованием нейросети SRCNN, применяемый для задачи восстановления текстур низкого разрешения.
2. провести экспериментальное исследование качества восстановленных текстур в зависимости от разрешения текстуры и применяемого сжатия.
3. сравнить получившиеся результаты – оценить размеры файлов, качество изображений визуально и с использованием метрики PSNR.

Алгоритм сжатия S3TC

S3 Texture Compression (S3TC, DXTn) – это семейство алгоритмов сжатия текстур, разработанных компанией S3 Graphics и широко используемых в различных графических приложениях, включая видеоигры и 3D-графику. DXT был создан для того, чтобы эффективно сжимать текстуры с потерями, снижая их размер, но при этом, стараясь минимизировать ухудшение качества изображения [3].

Особенностями данного алгоритма являются:

- Сжатие с потерями. DXT – после сжатия не все данные изображения можно восстановить. Однако сжатие осуществляется таким образом, чтобы визуальные искажения были минимальными для человеческого восприятия.

– Блочное сжатие. Алгоритм делит текстуры на блоки размером 4x4 пикселя, и каждый такой блок сжимается индивидуально. Это позволяет сжимать текстуры с высокой эффективностью, поскольку блоки могут содержать достаточно похожих пикселей, что позволяет эффективно кодировать информацию [4].

– Работа с цветом и альфа-каналом. DXT использует методы сжатия для хранения цветовой информации (RGB) и альфа-каналом (прозрачность), если таковой имеется.

Выбор данного алгоритма сжатия обусловлен тем, что он является более эффективным среди аналогов, за счет использования большего количества вычислительных ресурсов, а также поддержки альфа-канала. Это позволяет использовать его не только при разработке приложений для ПК, но и для мобильных устройств, в том числе VR-гарнитур.

Нейронная сеть SRCNN

Super-Resolution Convolutional Neural Network (SRCNN) представляет собой свёрточную нейронную сеть, предназначенную для восстановления изображения высокого разрешения из низкого за счёт обучения на паре «исходное изображение – его сжатая (или уменьшенная) версия» [2]. Метод опирается на принцип end-to-end обучения, при котором все стадии реконструкции, включая извлечение признаков и финальную сборку результата, осуществляются внутри единой архитектуры [2, 5]. Ранее в традиционных методах супер-разрешения, например на основе разреженного кодирования, каждая стадия решалась отдельно, и итоговое качество во многом зависело от оптимизации независимых этапов [2, 6].

SRCNN стал одной из первых глубоких моделей, продемонстрировавших результативность свёрточных сетей в задаче супер-разрешения [2, 7]. Сеть отличается компактной архитектурой, что обеспечивает высокую скорость восстановления изображения, а также гибкостью, позволяющей изменять количество слоёв или размер свёрточных фильтров для баланса между качеством и временем работы [2].

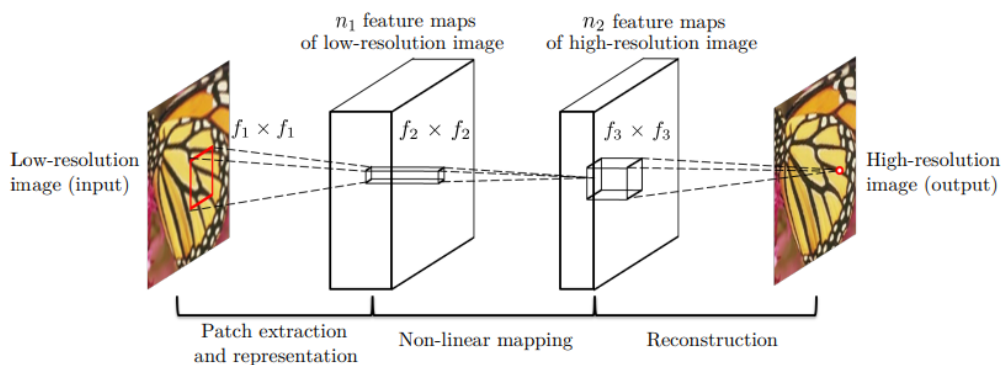


Рис. 1. Архитектура SRCNN

Тестирование различных моделей нейронной сети SRCNN

Для практической части было отобрано десять реализация нейросети SRCNN. В конечном итоге отсеяно было семь и тестирование осуществлялось на трёх моделях.

Первая модель использовала датасет Set5 и отличалась более гибкими настройками параметров супер-разрешения. При апскейле была возможность выбора конфигураций 9-1-5, 9-3-5, 9-5-5 и увеличения изображения в x2, x3, x4 раз.

Вторая модель использовала датасет Set5 с конфигурацией 9-5-5 и возможностью увеличения изображения в x2, x3, x4 раз.

Третья модель использовала датасет BSD100 с конфигурацией 9-1-5 и увеличением в x2, x3, x4 раз.

Датасет Set5 используется для оценки алгоритмов супер-разрешения изображений. Он состоит из 5 изображений различного содержания, которые широко применяются в качестве тестового набора для измерения качества работы алгоритмов повышения разрешения

изображений. Изображения датасета представлены в высоком и низком разрешениях. Низкое разрешение создается путем искусственного понижения разрешения, что позволяет оценивать работу алгоритмов по восстановлению деталей. Этот датасет часто используется для оценки качества работы алгоритмов супер-разрешения, таких как SRCNN, FSRCNN, EDSR, RCAN и для сравнения их эффективности по стандартным метрикам: PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index Measure).

Датасет BSD100 (Berkeley Segmentation Dataset 100) – набор изображений, используемый для оценки алгоритмов супер-разрешения изображений. Он широко применяется в научных исследованиях и статьях в качестве тестового датасета для нейросетей. Данный набор состоит из 100 изображений. Они взяты из более крупного набора BSD300, который был предназначен для задач сегментации изображений. По аналогии с Set5 данный набор содержит изображения высокого разрешения и соответствующие им версии низкого разрешения, полученные искусственным снижением разрешения при помощи бикубической понижающей дискретизации (bicubic downsampling).

Для тестирования моделей были выбраны изображения с различными разрешениями (256x256, 512x512, 1024x1024) и форматами (.png и .dds). Первая модель не обрабатывала сжатые DXT5 изображения формата .dds. В качестве альтернативы данные изображения были сконвертированы обратно в .png формат и использованы для тестов.

PSNR (Peak Signal-to-Noise Ratio) – это метрика, используемая для оценки качества изображения или видеосигнала после сжатия, передачи или другой обработки. Она показывает, насколько сильно изображение или видео было искажено в результате обработки по сравнению с оригиналом. Чем выше значение PSNR, тем лучше качество изображения и меньше искажений. Важно отметить, что PSNR не всегда точно отражает визуальное восприятие качества человеком. Например, даже при высоких значениях PSNR визуальный результат может казаться хуже, чем на изображении с более низким значением. Это зависит как от метода обработки, который применяется, так и от его реализации. На рисунке 2 показаны значения PSNR после увеличения изображений различными моделями. В названии изображений указано их оригинальное разрешение, масштаб увеличения, архитектура применённого супер-разрешения, конвертировано оно или нет. По данному графику можно сделать вывод о том, что модель №2 показывает наилучший результат.

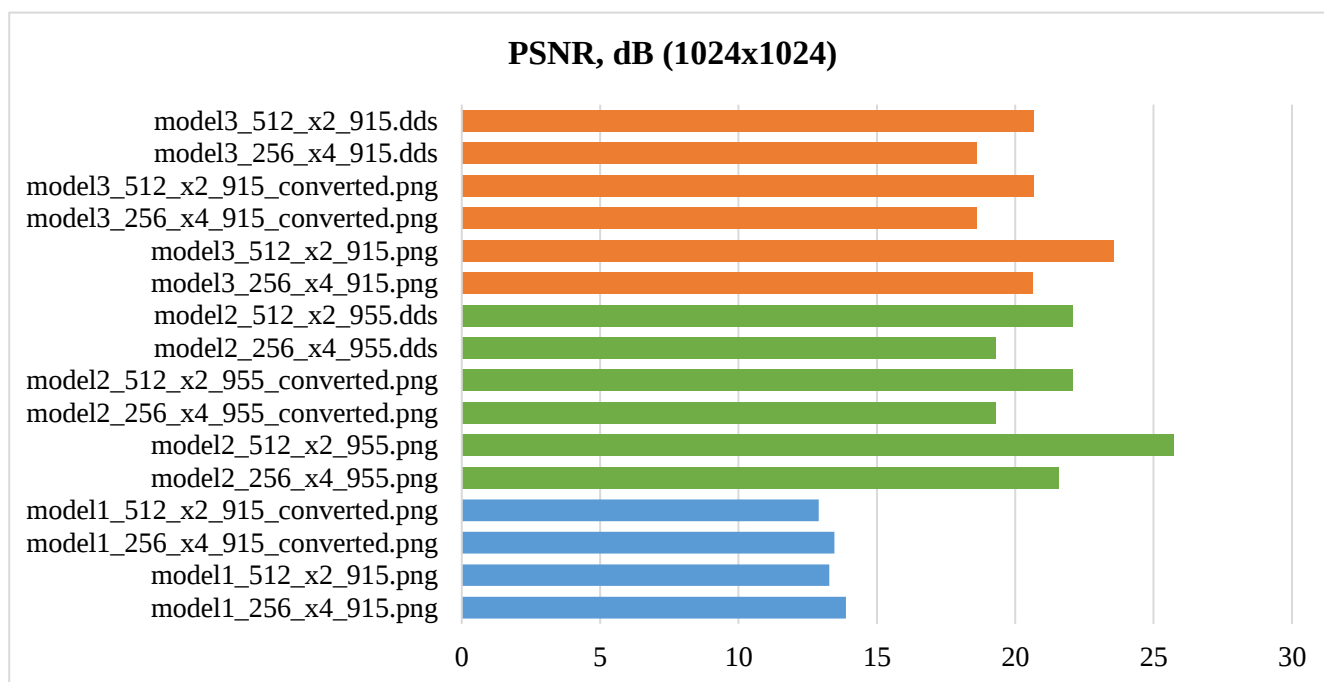


Рис. 2. Значения шума моделей при увеличении изображений до разрешения 1024x1024

В таблице 1 приведены размеры различных изображений и текстур. Если произвести оценку размеров и соответствующих значений PSNR, то можно сделать вывод о том, что обработка конвертированных изображений не даёт выигрыша в размере, но ухудшает выходное качество. Текстуры формата .dds полученные на вход в результате работы нейронной сети по значению PSNR равны конвертированным изображениям, но по размеру занимают больше места, чем оригинальное изображение 1024x1024.

Таблица 1. Размеры оригинальных изображений и обработанных с помощью различных моделей нейронной сети

Изображение	Размер, КБ	Сжатый, КБ	PSNR
original_512.png	699	342	
original_1024.png	2624	1366	
model1_256_x4_915.png	1540	501	13,885
model1_512_x2_915.png	2004	501	13,2773
model1_256_x4_915_converted.png	1564	-	13,4676
model1_512_x2_915_converted.png	2091	-	12,8959
model2_256_x4_955.png	1789	683	21,5796
model2_512_x2_955.png	2146	683	25,7283
model2_256_x4_955_converted.png	1856	-	19,2896
model2_512_x2_955_converted.png	2263	-	22,081
model2_256_x4_955.dds	3073	-	19,2896
model2_512_x2_955.dds	3073	-	22,081
model3_256_x4_915.png	1734	683	20,6339
model3_512_x2_915.png	2199	683	23,5424
model3_256_x4_915_converted.png	1762	-	18,5936
model3_512_x2_915_converted.png	2290	-	20,6598
model3_256_x4_915.dds	3073	-	18,5936
model3_512_x2_915.dds	3073	-	20,6598

Заметное увеличение размера .dds файла связано с тем, что алгоритм сжатия DXT5 хранит данные блоками по 4×4 пикселя, и каждый блок имеет фиксированный размер. Чем больше деталей появляется в изображении после работы нейронной сети SRCNN, тем менее эффективно DDS сжимает текстуру.

Таким образом, использование сжатого текстурного формата для супер-разрешения является нерациональным. Наиболее предпочтительный подход в данном случае в соотношении качество/вес файла – использовать для увеличения оригинальное .png изображение, которое затем сжимается при помощи DXT5 в текстуру .dds.

На рисунке 3 представлено сравнение выходных изображений с эталонным. Входные изображения с более высоким разрешением дают более качественные результаты на выходе. Модель №2 не смотря на лучшие показатели PSNR, справляется хуже с обработкой изображений низкого качества и имеет заметные артефакты, в отличие от модели №2. Модель №3 показала наихудший результат, в сравнении с двумя другими моделями. Несмотря на отсутствие заметных артефактов на выходных изображениях, они получились очень размытыми.



Рис. 3. Сравнение результатов работы различных моделей нейронной сети с эталонным изображением 1024x1024

Результаты

После тестирования трёх моделей можно сделать следующий вывод об алгоритме, который совмещает частичное сжатие DXT5 и работу сверточной нейронной сети SRCNN:

- модель №2 показывает наилучшее качество выходного изображения.
- модель №3 приближена по качеству выходного изображения к модели №2, но хуже обрабатывает входные изображения низкого разрешения и имеет заметные артефакты, несмотря на лучшие показатели по метрике PSNR.
- модель №1 обладает наибольшей гибкостью при настройке параметров для супер-разрешения, однако проигрывает в качестве выходного изображения моделям №1 и №2.
- обработка сжатых текстур .dds формата является нецелесообразной, так как размер выходного файла превышает размер эталонного несжатого изображения.
- обработка конвертированных изображений является нецелесообразной, так как качество входного изображения хуже оригинала и содержит артефакты, которые при супер-разрешении становятся еще более заметными.

– качество выходного изображения при использовании SRCNN прямо пропорционально разрешению входного изображения.

– использование оригинальных изображений для операций супер-разрешения с последующим сжатием алгоритма DXT5, является наиболее оптимальным вариантом с точки зрения отношения качество / размер файла.

– при необходимости можно сделать конвертацию .dds формата обратно в .png, это даст более лучший результат, чем использование конвертированного изображения для операции супер-разрешения.

Заключение

В результате выполнения данной работы была написана программа с использованием алгоритма сжатия DXT5, исследована сверточная нейронная сеть SRCNN, на практике рассмотрены десять моделей данной нейросети, из которых в качестве опытных образцов для тестов были выбраны три модели.

В ходе тестирования трёх моделей SRCNN была проверена работа увеличения для текстур разрешений 256x256, 512x512, 1024x1024 с заданными параметрами. Результаты работы были оценены по следующим характеристикам: значение PSNR, визуальное качество выходных изображений в сравнении с эталоном, размер файла.

После анализа результатов работы каждой из моделей были сделаны выводы и предложен алгоритм работы, совмещающий частичное сжатие и нейронную сеть SRCNN.

В качестве идей для дальнейшего исследования можно выделить следующие пункты:

– улучшение качества выходного изображения, которое можно получить, добавив дополнительный слой к SRCNN.

– улучшение качества выходного изображения за счет использования более новых моделей нейросетей для задач супер-разрешения.

– улучшение качества текстур за счет использования более новых алгоритмов сжатия.

Список использованных источников

1. Dong C., Loy C.C., He K., Tang X. Image Super-Resolution Using Deep Convolutional Networks // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2016. [Электронный ресурс] – URL: arxiv.org/pdf/1501.00092 (дата обращения: 02.03.2025).

2. Dong C., Loy C.C., He K., Tang X. Learning a Deep Convolutional Network for Image Super-Resolution // Proceedings of the European Conference on Computer Vision (ECCV). – 2014. [Электронный ресурс] – URL: sci-hub.ru/10.1007/978-3-319-10593-2_13?ysclid=m94i6y5zg7699018065 (дата обращения: 02.03.2025).

3. Сжатие текстур: S3TC (DXTn) // annimon [Электронный ресурс] – URL: annimon.com/article/984 (дата обращения: 21.12.2024).

4. Современные форматы сжатия текстур // sv-journal [Электронный ресурс] – URL: sv-journal.org/2014-1/06/ru/index.php?lang=en (дата обращения: 21.12.2024).

5. Burger H.C., Schuler C.J., Harmeling S. Image denoising: Can plain neural networks compete with BM3D? // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2012. [Электронный ресурс] – URL: hcburger.com/files/neuraldenoising.pdf (дата обращения: 10.03.2025).

6. Yang J., Wright J., Huang T., Ma Y. Image super-resolution via sparse representation // IEEE Transactions on Image Processing. – 2010. [Электронный ресурс] – URL: researchgate.net/publication/224138603_Image_Super-Resolution_Via_Sparse_Representation (дата обращения: 11.03.2025).

7. Freedman G., Fattal R. Image and video upscaling from local self-examples // ACM Transactions on Graphics. – 2011. [Электронный ресурс] – URL: cs.huji.ac.il/w~raananf/projects/lss_upscale/paper.pdf (дата обращения 12.03.2025).