

УДК 004.932.4

ИССЛЕДОВАНИЕ И РАЗРАБОТКА АЛГОРИТМОВ И ПРОГРАММНОЙ СИСТЕМЫ ДЛЯ УВЕЛИЧЕНИЯ РАЗРЕШЕНИЯ И УЛУЧШЕНИЯ КАЧЕСТВА ВИДЕОМАТЕРИАЛОВ

© Болотин Е.В., Сизов П.В.

e-mail: abcxyz@airmail.cc

*Рыбинский государственный авиационный технический университет
имени П.А. Соловьёва, г. Рыбинск, Российская Федерация*

В настоящее время уже существуют классические алгоритмы для увеличения изображений (например, билинейная или бикубическая интерполяция). Однако они не способны в полной мере восполнить ту информацию, которая была потеряна при сжатии или конвертировании изображения из одного формата в другой. Так, например, алгоритм JPEG сильно подвержен потере информации. Всё происходит потому, что JPEG является форматом сжатия с потерями, и при каждом сохранении растёт количество математических усреднений, ошибок. Два сохранения с 90 % сжатием примерно эквивалентны одному с 81 % по количеству артефактов. Точно такая же ситуация сложилась и в отношении видеоматериалов. Для экономии пространства накопителей видеоматериалы подвергаются сжатию с потерями и уменьшению разрешения. Это приводит к появлению и распространению низкокачественного видеоконтента. Кроме того, существует и множество других вариантов улучшения изображений, например, шумоподавление или регулировка цвета и контраста, которые всё равно не позволяют восстановить потерянную информацию. Для спутниковых снимков одним из показателей качества является Ground Sample Distance (GSD), физическое измерение, представляемое одним пикселем на изображении. Для улучшения таких изображений применяют технологию, называемую «суперразрешением».

Технология «суперразрешения» (SR) выполняет задачу улучшения некачественных и сжатых изображений высокого разрешения из более мелких изображений (синтезируется субпиксельная информация небольших изображений для обработки основного изображения) и имеет очень важное значение для применения в областях анализа спутниковых и медицинских изображений [1].

Синтез включают в себя следующие методы:

- интерполяция соседних пикселей в изображении;
- интерполяция близлежащих кадров в видео;
- частотная фильтрация для уменьшения шума.

В рамках данной работы планируется изучить применяемые методы в направлении использования искусственных нейронных сетей.

Для выполнения поставленной цели предлагается исследовать модель SRCNN, которая реализуется в следующем алгоритме:

- 1) выполнение интерполяции для увеличения изображения с низким разрешением до необходимого размера;
- 2) формирование нелинейной карты через трехслойную свёрточную сеть;
- 3) восстановление изображения с высоким разрешением.

Как правило, для первичного увеличения изображения используется алгоритм бикубической интерполяции.

Обозначим как Y (рис.) входное изображение, которое получено в результате интерполяции, а изображение, с которым будет осуществляться сравнение, как X . В ходе работы необходимо получить изображение $F(Y)$, которое должно быть сравнимо (максимально схоже) по качеству с изображением X . Для этого нужно найти функцию F , выполнив следующие шаги [2]:

1) **извлечение патчей** операция позволяет получить участки изображения (патчи), состоящие из нескольких пикселей, из изображения с низким разрешением Y . Затем необходимо представить каждый такой патч в виде набора карт признаков, размер которых равен размерности вектора;

2) **нелинейное отображение** операция позволяет получить нелинейное отображение каждого многомерного вектора на другие векторы. Каждый отображенный вектор (часть многомерного вектора) представляет патч высокого разрешения уже с другим набором карт признаков;

3) **восстановление** на данном шаге воссоздается изображение $F(Y)$, для чего используются отображения высокого разрешения каждого патча, полученные в предыдущем шаге. Ожидается, что в результате полученное изображение будет схожим с исходным изображением X .

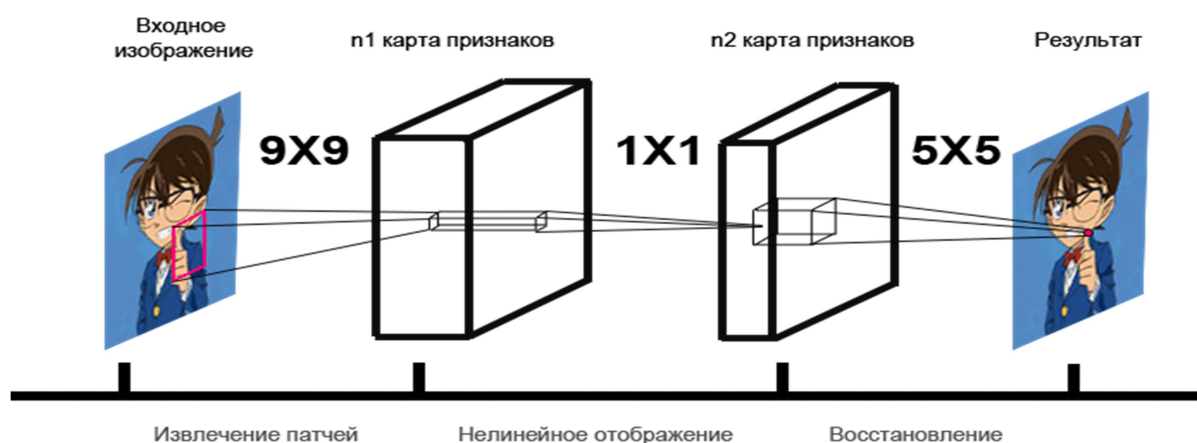


Рис. модель SRCNN

Всего данная модель содержит три слоя:

1. Первый слой.

Используется для получения патчей и обозначается формулой:

$$F_1(Y) = \max(0, W_1 * Y + B_1),$$

где W_1 – фильтры, B_1 – веса, * – операция свёртки.

Фильтры W_1 можно обозначить через формулу:

$$W_1 = c \times f1 \times f1,$$

где c – число каналов изображения Y , $f1$ – пространственный размер фильтра.

Фильтры W_1 осуществляют $n1$ операций свёртки изображения, и каждая свёртка имеет ядро свёртки размера $c \times f1 \times f1$. Выходные данные слоя содержат $n1$ карт признаков.

Веса B_1 представляют собой $n1$ -мерный вектор, каждый элемент которого сопоставлен с элементом фильтра W_1 .

В качестве активационной функции первого слоя используется ReLU (Rectified Linear Unit)

$$f(x) = \max(0, x),$$

где x – входной сигнал.

2. Второй слой.

Слой производит операцию нелинейного отображения $n1$ -мерных векторов на $n2$ -мерные векторы. Это будет работать только для фильтров с размером 1×1 . Описанная операция выражается формулой:

$$F_2(Y) = \max(0, W_2 * F_1(Y) + B_2),$$

где параметры W_2 и B_2 аналогичны подобным параметрам для F_1 с той разницей, что теперь фильтры осуществляют $n2$ операций, а веса представлены $n2$ -мерным вектором. Так же, как и у первого слоя используется ReLU.

3. Третий слой.

Слой осуществляет финальную операцию восстановления изображения. Его можно представить следующей формулой:

$$F(Y) = W_3 * F_2(Y) + B_3,$$

где W_3 – это $c \times c$ фильтров размерностью $n2 \times f3 \times f3$, B_3 – это $n2$ -мерный вектор. Выходными данными этого слоя является изображение высокого разрешения. Для активации данного слоя используется линейная функция:

$$f(x) = x,$$

где x – это входной сигнал.

В качестве функции потерь используется среднеквадратическая ошибка MSE (Mean Squared Error):

$$L(\theta) = 1/n \sum_{i=1}^n |F(Y_i) - X_i|^2,$$

где θ – это конфигурация нейронной сети $W_1, W_2, W_3, B_1, B_2, B_3$, n – число сэмплов обучающей выборки [2].

Улучшение качества видео подразумевает несколько дополнительных операций по разложению на кадры и обратную сборку видеодорожки. В этом случае количество данных для обработки будет существенно увеличено по сравнению с простой обработкой серии изображений. Это значит, что для улучшения одного видеофайла в разумные сроки необходимо будет использовать графические процессоры для ускорения процесса. Получить кадры из видеодорожки можно, если воспользуемся библиотекой ffmpeg:

```
ffmpeg -i video.raw image%08d.bmp,
```

где video.raw – исходная видеодорожка, image%08d.bmp – выходной кадр в формате BMP.

Для обучения модели могут использоваться различные наборы изображений. Однако каждое переобучение может как улучшить, так и ухудшить выходное качество для данной модели, поэтому следует обоснованно выбирать наборы данных под каждую задачу.

Ближайшими целями работы являются: программная реализация представленного способа для задачи восстановления разрешения видеоматериалов и определение направлений его развития.

Библиографический список

1. Learning a Deep Convolutional Network for Image Super-Resolution [Электронный ресурс]. URL: http://personal.ie.cuhk.edu.hk/~ccloy/files/eccv_2014_deepresolution.pdf
2. Image Super-Resolution Using Deep Convolutional Networks [Электронный ресурс]. URL: <https://arxiv.org/pdf/1501.00092.pdf>