

НЕЙРОННЫЕ СЕТИ: МОДИФИКАЦИЯ АЛГОРИТМА СЖАТИЯ ИЗОБРАЖЕНИЙ

Трофимук В.А., Цветков А.А.

Белорусский государственный университет, г. Минск

В качестве базовой была взята методика сжатия, применявшаяся при обработке изображений спектрометрической системы орбитальной космической станции «МИР». Данный алгоритм был модифицирован.

Основные моменты базового алгоритма [1].

Исходное двумерное изображение представляется в виде совокупности $x(k) = (x_1(k), \dots, x_m(k))$, $k = 1, \dots, N$, $m = \mu_1 \mu_2$ непересекающихся фрагментов. Фрагменты представляют собой образцы, предъявляемые сети для обучения.

Возьмём двухслойную сеть. Первый слой содержит n нейронов. Его вход — вектор $\mathbf{p} = \mathbf{x}(k)$, а выход — вектор $\mathbf{a}^{(1)}$ с компонентами

$$a_i^{(1)}(k) = \sum_{j=1}^m w_{ij}^{(1)} x_j(k), \quad i = 1..n$$

Этот слой является кодером. Для каждого фрагмента $\mathbf{x}(k)$ вычисляется вектор коэффициентов кодирования $\mathbf{a}^{(1)}(k)$.

Роль декодера выполняет второй слой сети, который по вектору $\mathbf{a}^{(1)}(k)$ пытается восстановить фрагмент $\mathbf{x}(k)$:

$$a_j^{(2)}(k) = \sum_{i=1}^n w_{ji}^{(2)} a_i(k), \quad j = 1..m$$

Значения коэффициентов весовых матриц $\mathbf{W}^{(1)}$ и $\mathbf{W}^{(2)}$ не зависят от фрагментов и определяются условием достижения минимума ошибки

$$\min_{w^{(1)}, w^{(2)}} \hat{F}, \quad \hat{F} = \sum_{j=1}^m \left[x_j(k) - a_j^{(2)}(k) \right]^2$$

Для определения весовых коэффициентов применяется базовый вариант алгоритма обратного распространения ошибки [2].

Для восстановления изображения нужно сохранить все $x_j(k)$ и параметры декодера $\mathbf{W}^{(2)}$. Из-за заголовка с параметрами реальный коэффициент сжатия меньше m/n , он определяется не только количеством нейронов в каждом из слоев, но и размером изображения, и при увеличении размера асимптотически приближается к m/n (качество при этом ухудшается). Если число точек изображения равно P , то выражение для коэффициента сжатия имеет вид:

$$\gamma = \frac{m}{n} \frac{1}{1 + \frac{m^2}{P}}$$

Предложенная нами модификация алгоритма заключается в следующем. Изображение делится на l независимых сегментов, которые сжимаются описанным выше способом, и в архив записываются l наборов $\mathbf{W}^{(2)}$. Коэффициент сжатия для конкретного изображения определяется его размером и значениями l , m и n . Одному и тому же коэффициенту сжатия соответствуют различные тройки (l, m, n) , качество при этом разное. Выражение для коэффициента сжатия примет вид:

$$\gamma = \frac{m}{n} \frac{1}{1 + \frac{m^2 l}{P}}$$

Смысл использования модифицированного алгоритма становится понятен, если учесть, что внутри небольшого фрагмента цвета точек близки, а сами фрагменты, находящиеся в разных частях изображения, существенно отличаются. Так, пусть требуется сжать в два раза достаточно большое изображение, содержащее сегмент, описываемый восемнадцатью числами от 0 до 34 через 2. Если применять базовый алгоритм, то можно взять $m=2$, $n=1$. Изображение велико, за счет чего, во-первых, фактический коэффициент сжатия достаточно точно равен двум, во-вторых, из-за совершенно различных цветов в разных частях изображения сжатие сведется к замене чисел, задающих цвета пары соседних точек, на их среднее. При этом средняя ошибка передачи указанного сегмента составит 1 тчк⁻¹. Если использовать модифицированный алгоритм с $(m, n)=(3, 1)$ и восемнадцатью точками в сегменте, то, как показывает несложный расчет, средняя ошибка передачи указанного сегмента составит 0,71 тчк⁻¹, или, с учетом округлений, 0,67 тчк⁻¹.

Литература

1. Голубев Ю.Ф. Нейронные сети в мехатронике //Фундаментальная и прикладная математика. - М.: МГУ, 2005, Т. 11, Вып. 8, С. 81-103
2. Лутковский В.М. Нейронные сети: Конспект лекций. - Мн.: БГУ, 2003.