

А.А. СИРОТА, А.В. АКИМОВ, Р.Р. ОТЫРБА  
**ДЕФОРМИРУЮЩИЕ ПРЕОБРАЗОВАНИЯ ИЗОБРАЖЕНИЙ  
И ИХ ПРИМЕНЕНИЕ ПРИ АУГМЕНТАЦИИ ДАННЫХ  
ДЛЯ ОБУЧЕНИЯ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ**

*Сирота А.А., Акимов А.В., Отырба Р.Р. Деформирующие преобразования изображений и их применение при аугментации данных для обучения глубоких нейронных сетей.*

**Аннотация.** Проведены исследования возможностей аугментации (искусственного размножения) обучающих данных в задаче классификации с использованием деформирующих преобразований обрабатываемых изображений. Представлены математическая модель и быстродействующий алгоритм выполнения деформирующего преобразования изображения, при использовании которых исходное изображение преобразуется с сохранением своей структурной основы и отсутствием краевых эффектов. Предложенный алгоритм используется для аугментации наборов изображений в задаче классификации, содержащих относительно небольшое количество обучающих примеров. Аугментация исходной выборки осуществляется в два этапа, включающих зеркальное отображение и деформирующее преобразование каждого исходного изображения. Для проверки эффективности подобной техники аугментации в статье проводится обучение нейронных сетей – классификаторов различного вида: сверточных сетей стандартной архитектуры (convolutional neural network, CNN) и сетей с остаточными связями (deep residual network, DRN). Особенностью реализуемого подхода при решении рассматриваемой задачи является также отказ от использования предобученных нейронных сетей с большим количеством слоев и дальнейшим переносом обучения, поскольку их применение несет за собой затраты с точки зрения используемого вычислительного ресурса. Показано, что эффективность классификации изображений при реализации предложенного метода аугментации обучающих данных на выборках малого и среднего объема повышается до статистически значимых значений используемой метрики.

**Ключевые слова:** глубокие нейронные сети, аугментация обучающих данных, деформирующие искажения изображений, эффективность глубоких нейронных сетей.

**1. Введение.** В современных обучающихся системах зачастую возникает проблема недостатка данных, используемых для обучения, связанная как с ресурсными ограничениями при их подготовке, так и с объективными факторами, мешающими получить достаточный набор примеров. Для разрешения этой проблемы в системах классификации на основе глубоких нейронных сетей (ГНС) применяются разнообразные технологии аугментации данных, используемых для обучения. Под аугментацией далее будем понимать искусственное размножение данных (ИРД), т.е. использование некоторых «опорных» образов для увеличения объема обучающих выборок на основе стохастических или детерминистских моделей представления данных. Применение аугментации в условиях малой и средней по объему выборки во многих случаях позволяет получить более высокую

эффективность алгоритмов классификации и сегментации изображений при их тестировании на новых данных. Такой подход к подготовке обучающих данных позволяет снизить затраты времени, а также, в определенной степени, преодолеть проблему несбалансированности обучающих данных различных классов.

Существует и альтернативный подход в случаях, когда сбор необходимого числа обучающих образов оказывается сложен из-за специфического характера предметной области. Он состоит в использовании заранее предобученных моделей классификаторов. Так, часто реализуемым подходом применительно к практике применения ГНС является использование готовых (предобученных) нейронных сетей с большим количеством слоев и последующей реализации техники переноса обучения в контексте решаемой задачи на основе имеющейся малой выборки, которая применяется для дообучения. Однако такой подход видится не всегда оправданным, так как использование архитектуры предобученных сетей требует существенных вычислительных ресурсов. Поэтому вариант с проведением обучения сетей собственной архитектуры, удовлетворяющей требованиям по ресурсоемкости, остается востребованным.

Вопросы повышения качества классификации образов, основанные на ИРД, продемонстрированы в работах [1–15]. В большинстве из них для генерации новых изображений в обучающей выборке при обучении ГНС используются различные преобразования: поворот, сжатие и растяжение, наклон, зеркальное отражение, обрезка, смещение и другие. Возможности современных программных сред, например, Keras (Tensorflow) предусматривают включение опций по использованию подобных стандартных средств аугментации. В работах отечественных авторов [7–15] также представлено применение различных сочетаний алгоритмов аугментации в задачах обработки изображений.

В [7] новые данные в обучающей выборке генерируются с помощью морфинг-преобразований путем «скрещивания» исходных данных между собой. В [8] предложены алгоритмы внесения реалистической деформации изображений лиц, с помощью которых была размножена стандартная обучающая выборка изображений для алгоритма Виолы-Джонса. Показано, что подобным образом объем обучающей выборки может быть уменьшен в примерно в 10 раз при снижении вероятности распознавания не более чем на 2–4%. В [10] описывается алгоритм ИРД для задач машинного обучения классификаторов биологических объектов по спектрам, основанный

на использовании ядерных оценок функций правдоподобия классов образов.

Существуют также реализации алгоритмов деформирующих преобразований, используемых при аугментации изображений, включенные в современные библиотеки и фреймворки для Python, такие, например, как библиотека Albumentations. Наиболее популярный среди данной категории алгоритм, известный под названием ElasticTransform [16], в Albumentations ориентирован на аугментацию при решении задач классификации, сегментации и поиска объектов на изображениях и основан на использовании разнообразных методов интерполяции данных. Примеры применения этого алгоритма представлены в [15, 17, 18].

Помимо задач аугментации данных [4, 5, 8, 9] в известных работах описано использование деформирующих преобразований и их моделей внутри самих алгоритмов распознавания [19–21], для генерации и модификации изображений при построении сверхразрешения [22], 3D-моделировании [9, 23], для защиты биометрических систем от атак с использованием этих деформаций [24, 25]. В [26, 27] представлены реализации данных подходов на основе ГНС.

В целом по результатам ранее выполненных исследований следует отметить высокую эффективность применения деформирующих преобразований изображений, показанную при обучении алгоритмов распознавания.

В то же время, следует отметить, что применение любых алгоритмов аугментации сопряжено с увеличением времени обучения, вследствие естественного его расхода на преобразование входных изображений. В особенности это касается алгоритмов, реализующих деформирующие преобразования, как весьма затратных.

В связи с изложенным целью настоящей работы является исследование вопросов построения быстродействующих алгоритмов формирования плавных деформирующих преобразований (ДП), обеспечивающих сохранение структурной основы исходного изображения, а также возможностей их применения для аугментации данных при обучении ГНС различной архитектуры в задачах классификации и оценки получаемого при этом выигрыша.

**2. Математическая модель и алгоритм деформирующих преобразований.** Пусть  $\Omega_x \subset \mathbb{R}^n$  – некоторое множество значений аргумента непрерывной функции. Определим деформирующее преобразование (ДП) скалярной функции  $n$  переменных

$f: \Omega_x \rightarrow G \subset \mathbb{R}^1$ ,  $\mathbf{g} = f(\mathbf{x})$ ,  $\mathbf{x} \in \Omega_x$ ,  $\mathbf{g} \in G$  как преобразование, модифицирующее значение ее аргументов в области определения по заданному закону. Очевидным способом такой модификации является добавление значения функции деформации известного вида к исходным значениям этих переменных по следующей формуле:

$$\begin{aligned} f[x_1 + r_1(x_1, \dots, x_n), \dots, x_n + r_n(x_1, \dots, x_n)] = \\ = f[u_1(x_1, \dots, x_n), \dots, u_n(x_1, \dots, x_n)] = g(x_1, \dots, x_n), \end{aligned} \quad (1)$$

где  $g(x_1, \dots, x_n)$  – результирующая деформированная функция;  $r_i(x_1, \dots, x_n)$ ,  $i = \overline{1, n}$  – непрерывные функции ДП по каждой координате, которые могут быть детерминированными функциями фиксированной формы или представлять реализации многомерного случайного поля.

Перейдем к использованию векторных обозначений  $\mathbf{x} = (x_1, \dots, x_n)^T$ ,  $\mathbf{r}(\mathbf{x}) = (r_1(\mathbf{x}), \dots, r_n(\mathbf{x}))^T$  и  $\mathbf{u}(\mathbf{x}) = (u_1(\mathbf{x}), \dots, u_n(\mathbf{x}))^T$ . Тогда можно записать выражение для выполняемого ДП в векторном виде следующем образом:

$$f[\mathbf{x} + \mathbf{r}(\mathbf{x})] = f[\mathbf{u}(\mathbf{x})] = g(\mathbf{x}).$$

При реализации алгоритмов ДП на основе (1) следует учесть ряд особенностей реализуемых преобразований.

Первая из них заключается в том, что для корректного формирования значений деформированной функции  $f(\mathbf{u}(\mathbf{x})) = g(\mathbf{x})$ , значения ее аргументов не должны выходить за пределы области определения исходной функции  $f(\mathbf{x})$ , имеющей в общем случае вид многомерной регулярной дискретной сетки. Другими словами, если для  $f(\mathbf{x})$  вектор  $\mathbf{x} \in \Omega_x$ , то необходимо, чтобы вектор  $\mathbf{u}(\mathbf{x}) \in \Omega_x$ .

Подобное условие может быть выполнено [28], если использовать специальные масочные функции  $\mathbf{h}(\mathbf{x}) = (h(x_1), \dots, h(x_n))^T$ , ограничивающие краевые значения  $\mathbf{u}(\mathbf{x})$ :

$$\mathbf{u}(\mathbf{x}) = \mathbf{x} + \mathbf{u}'(\mathbf{x}) = \mathbf{x} + \mathbf{h}(\mathbf{x}) \otimes \mathbf{r}(\mathbf{x}),$$

где  $\otimes$  – обозначение операции поэлементного умножения векторов. В качестве масочных функций могут быть использованы оконные

функции, имеющие трапецевидную форму или форму многомерного гауссовского закона распределения. В дальнейшем в реализованных алгоритмах обработки изображений будут использоваться трапецевидные функции, которые, в отличие от гауссиан, практически не модифицируют функцию деформирующего преобразования в центральной области определения.

Вторая особенность алгоритмов внесения ДП состоит (по крайней мере, для задач обработки изображений) в необходимости их применения по отношению к решетчатым функциям, заданным на многомерных дискретных сетках. В этом случае, как показано в [8] для обеспечения возможности модификации аргументов решетчатой функции  $\hat{f}(\mathbf{x})$  при внесении в них непрерывных деформаций произвольного характера необходимо предварительно выполнить ее интерполяцию и представление в виде функции  $\tilde{f}(\mathbf{x})$  вещественных переменных. Затем для интерполированной непрерывной функции необходимо выполнить обратный переход к дискретному представлению деформированной решетчатой функции  $\hat{g}(\mathbf{x})$ , путем дискретизации с исходным шагом.

Анализ применения возможных способов интерполяции, например, с использованием радиально-базисных функций или сплайна тонкой пластины показывает, что при реализации подобных подходов, оптимизированные алгоритмы имеют сложность порядка  $O(N_{xy} \cdot \log(N_{xy}))$  [29–31], где  $N_{xy}$  – общее число элементов интерполируемого массива данных произвольной размерности (в данном случае общее число пикселей изображения). Также при использовании интерполяции дополнительные накладные расходы возникают при выполнении сдвиговых деформаций, сложность которых при заранее вычисленной функции деформации можно оценить как  $O(N_{xy})$ . Все это делает подобные алгоритмы ДП затратными в вычислительном отношении, особенно, если требуется преобразовать большое число изображений при аугментации обучающих примеров. Поэтому предлагается реализовать упрощенный подход, основанный на использовании ограничивающей маски и реализации перестановки элементов решетчатой функции в соответствие с используемым механизмом внесения ДП, сложность которого составляет  $O(N_{xy})$ .

Рассмотрим применительно к задаче обработки изображений следующую постановку. Пусть дифференцируемая функция

$f(x_1, x_2)$ ,  $\mathbf{x}=(x_1, x_2) \in \Omega_x$  определена на прямоугольной области  $\Omega_x$  размера  $R_1 \times R_2$ . Требуется преобразовать ее путем внесения ДП:

$$g(x_1, x_2) = f[x_1 + A_m r_1(x_1, x_2), x_2 + A_m r_2(x_1, x_2)],$$

где  $\mathbf{r}(\mathbf{x}) = (r(x_1, x_2), r(x_1, x_2))^T$  – непрерывные функции, описывающие деформирующие искажения и являющиеся либо функциями известной формы, либо, в общем случае, реализациями случайного поля, причем  $|r_j(\mathbf{x})| \leq 1, j=1, 2$ ;  $A_m$  – амплитуда ДП, обеспечивающая с учетом ограничений, вносимых наложением масочной функции, выполнение условия  $\mathbf{x} + \mathbf{u}'(\mathbf{x}) \in \Omega_x$ .

При цифровой обработке изображений в результате дискретизации по пространственным координатам с интервалом  $\delta x = \delta x_1 = \delta x_2$  формируется цифровой эквивалент исходной функции путем фиксации ее значений в точках  $\hat{f}(i, k)$  и цифровой эквивалент  $\hat{g}(i, k)$  деформированной функции, заданных на прямоугольной дискретной сетке  $(i, k) \in \Psi = \{i = \overline{1, N}, k = \overline{1, M}\}$ :

$$\begin{aligned} \hat{f}(i, k) &= f(t_i, s_k), \\ t_i &= (i-1)\delta x, i = \overline{1, N}, s_k = (k-1)\delta x, k = \overline{1, M}, \\ N &= R_1 / \delta x, t_1 = 0, t_N = R_1, M = R_2 / \delta x, s_1 = 0, s_M = R_2, \\ \hat{g}(i, k) &= g(t_i, s_k) = f[t_i + A_m r_1(t_i, s_k), s_k + A_m r_2(t_i, s_k)]. \end{aligned} \quad (2)$$

Введем вектор целочисленных индексов  $\mathbf{v}(i, k) = (v_1(i, k), v_2(i, k))^T$ ,  $v_1(i, k) = \text{round}[A_m r_1(t_i, s_k) / \delta x]$ ,  $v_2(i, k) = \text{round}[A_m r_2(t_i, s_k) / \delta x]$ , где  $\text{round}[\dots]$  обозначает операцию округления до ближайшего значения. Тогда компоненты аргумента в (2) можно представить как:

$$\begin{aligned} t_i + A_m r_1(t_i, s_k) &= \delta x(i-1) + \delta x v_1(i, k) + \varepsilon_1(i, k), \\ s_k + A_m r_2(t_i, s_k) &= \delta x(k-1) + \delta x v_2(i, k) + \varepsilon_2(i, k), \\ -\delta x / 2 \leq \varepsilon_1(i, k) &< \delta x / 2, -\delta x / 2 \leq \varepsilon_2(i, k) < \delta x / 2. \end{aligned} \quad (3)$$

Вектор  $\boldsymbol{\varepsilon}(i, k) = (\varepsilon_1(i, k), \varepsilon_2(i, k))^T$  в (3) также является случайным. Он характеризует погрешность квантования по уровню

заданной в непрерывном времени функции деформации, которая сопутствует процессу дискретизации исходного изображения по пространству:

$$\begin{aligned} \hat{g}(i, k) &= g(t_i, s_k) = \\ &= f[t_i + \delta x v_1(i, k) + \varepsilon_1(i, k), s_k + \delta x v_2(i, k) + \varepsilon_2(i, k)]. \end{aligned} \quad (4)$$

На основе представления ДП в (4) используем для цифрового деформированного изображения первое приближение:

$$\begin{aligned} \hat{g}(i, k) &= f[\delta x(i-1) + \delta x v_1(i, k), \delta x(k-1) + \delta x v_2(i, k)] + \\ &+ f'_t[t_i + \delta x v_1(i, k), s_k + \delta x v_2(i, k)] \varepsilon_1(i, k) + \\ &+ f'_s[t_i + \delta x v_1(i, k), s_k + \delta x v_2(i, k)] \varepsilon_2(i, k), \end{aligned}$$

где  $f'_t, f'_s$  – частные производные исходной функции  $f$ . Тогда для (4) можно записать в скалярной и операторной форме:

$$\begin{aligned} \hat{g}(i, k) &= \hat{f}(i + v_1(i, k), k + v_2(i, k)) + O(\varepsilon_1) + O(\varepsilon_2), \\ \hat{\mathbf{g}} &= G[\hat{\mathbf{f}}, \mathbf{v}] + G[\hat{\mathbf{f}}'_t, \mathbf{v}] \otimes \varepsilon_1 + G[\hat{\mathbf{f}}'_s, \mathbf{v}] \otimes \varepsilon_2, \end{aligned} \quad (5)$$

где  $\hat{\mathbf{g}}$  – обозначение полученной после перестановки решетчатой функции;  $G[\hat{\mathbf{f}}, \mathbf{v}]$  – оператор перестановки позиций размещения исходной решетчатой функции  $\hat{f}(i, k)$  в соответствии со значениями  $\mathbf{v}(i, k)$ ;  $G[\hat{\mathbf{f}}'_t, \mathbf{v}]$  – обозначение синхронно выполняемых операторов перестановки производных функции  $f$ , взятых в точках дискретизации по пространству.

Оператор  $G(\hat{\mathbf{f}}, \mathbf{v})$  является случайным оператором, определяющим возможные размещения элементов  $\hat{\mathbf{f}}$  и  $\hat{\mathbf{f}}'$  в элементы  $\hat{\mathbf{g}}_1 = G[\hat{\mathbf{f}}, \mathbf{v}]$  и  $\hat{\mathbf{g}}_2 = G[\hat{\mathbf{f}}', \mathbf{v}]$ , соответственно. Для случайной функции ДП он задает вероятностные переходы при преобразовании множества пикселей исходного изображения  $\hat{\mathbf{f}}$  во множество пикселей изображения  $\hat{\mathbf{g}}$ .

Оператор является сюръекцией, в том смысле, что каждый элемент деформированной функции, является образом хотя бы одного

элемента исходной. Обратное, вообще говоря, неверно. Как уже сказано, возможны незначительные выпадения пикселей исходного изображения и их замещение близко расположенными пикселями этого же изображения. Также при размещении возможны повторения пикселей исходного изображения в случае, если изменения деформации происходят с отрицательной производной.

Анализ показывает, что значения  $\epsilon(i, k)$  в различных точках пространства малы (сопоставимы с погрешностью квантования исходной функцией деформации) и практически не коррелированы, что позволяет считать ее воздействие близким к воздействию белого шума с существенно меньшей относительно первого слагаемого амплитудой. С точки зрения решения задачи аугментации изображений это влияние будет незначительным и, скорее всего, иметь позитивное значение, так как дополнительное зашумление часто используется в стандартных схемах аугментации данных.

В ходе дальнейших исследований на основе рассмотренной математической модели были реализованы алгоритмы внесения ДП, имеющие воздействие в виде биполярных импульсов квазидетерминированной формы.

Для исключения краевых эффектов и ограничения ДП использовалась трапецевидная масочная функция, задаваемая в виде матрицы размера  $N \times M$  и однозначно исключаяющая краевые эффекты:

$$\begin{aligned}
 h(i, k) &= \tilde{h}_1(i)\tilde{h}_2(k), \quad \tilde{h}_1(i) = \frac{h_1(i)}{\max[h_1(i)]}, \\
 \tilde{h}_2(k) &= \frac{h_2(k)}{\max[h_2(k)]}, \quad i = \overline{1, N}, \quad k = \overline{1, M}, \\
 h_1(i) &= \begin{cases} i-1, & 1 \leq i \leq i_x+1, \\ i_x, & i_x+1 < i \leq N-i_x-1, \\ N-i-1, & N-i_x \leq i \leq N. \end{cases} \quad h_2(k) = \begin{cases} k-1, & 1 \leq k \leq k_y+1, \\ k_y, & k_y+1 < k \leq M-k_y-1, \\ M-k-1, & M-k_y \leq k \leq M. \end{cases}
 \end{aligned} \tag{6}$$

где  $i_x = k_y = 2A_m - 1$  – целочисленное (в ед. пикселей) представление диапазона возможных значений деформаций, определяющее расстояние от края области определения деформируемой функции, на котором начинают действовать ограничения. Значение  $A_m$  при этом фактически определяет максимальное количество позиций, на которое



может быть переставлен один элемент. Для всех граничных элементов значение маски равно нулю. В целом, анализ формы функции (6) показывает, что ни один из элементов исходного изображения при перестановке не выйдет за границы области определения. Такая масочная функция уже может быть использована в качестве функции деформации, если мы хотим ограничиться кусочно-линейным характером внесенных искажений типа сдвига и растяжения.

Для иллюстрации на рисунке 1 представлен граф переходов, описывающий перестановку элементов вдоль одной координаты (аналогичную по сути одномерной перестановке элементов последовательности) в соответствии с функцией ДП, заданной только на основе ограничивающей маски вида  $h_1$ . Количество переставляемых элементов равно 20. Вид функции, обеспечивающей такие перестановки, представлен штриховой линией. Здесь мы видим, что краевые элементы не меняются, часть элементов выпадает, а часть замещает выпадающие с повторениями.

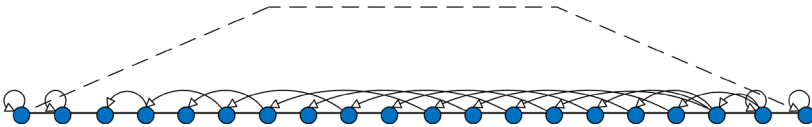


Рис. 1. Граф переходов, описывающий перестановку элементов вдоль одной координаты в соответствии с функцией ДП на основе маски вида  $h_1$

На рисунке 2(а) представлен вид получаемой таким образом функции деформации, заданной  $h(i, k) = \tilde{h}_1(i)$ ,  $i = \overline{1, N}$  одинаковым образом для каждой координаты с индексом  $k = \overline{1, M}$  ( $h_2(k) = 1$ ,  $i = \overline{1, M}$ ) при  $N = M = 200$ ,  $A_m = 25$ . На рисунке 2(в) представлено эталонное изображение в виде регулярной шахматной структуры  $N \times M$ ,  $N = M = 200$ , а на рисунке 2(г) – его преобразование в соответствии с используемой таким образом функцией деформации.

Анализ рисунка 2(а, в, г) показывает, что здесь смещение элементов осуществляется только в одном направлении. В результате такого ДП, примененного к изображению, происходит «дружный» сдвиг элементов, размещенных по вертикали, в левой части изображения, и растяжение элементов, также размещенных по вертикали, в правой части изображения. Такое преобразование, по сути, соответствует схеме перестановки, представленной на рисунке 1.

Как уже отмечалось, при использовании соотношений для  $h_1(i)$ ,  $i = \overline{1, N}$ ,  $h_2(k)$ ,  $k = \overline{1, M}$  из (6) в полной мере в результате получается ограничивающая по обеим координатам маска трапецевидной формы, которая и будет далее использована во всех примерах и анализируемых моделях. Отображение ее формы для  $N = M = 200$ ,  $A_m = 25$  представлено на рисунке 2(б).

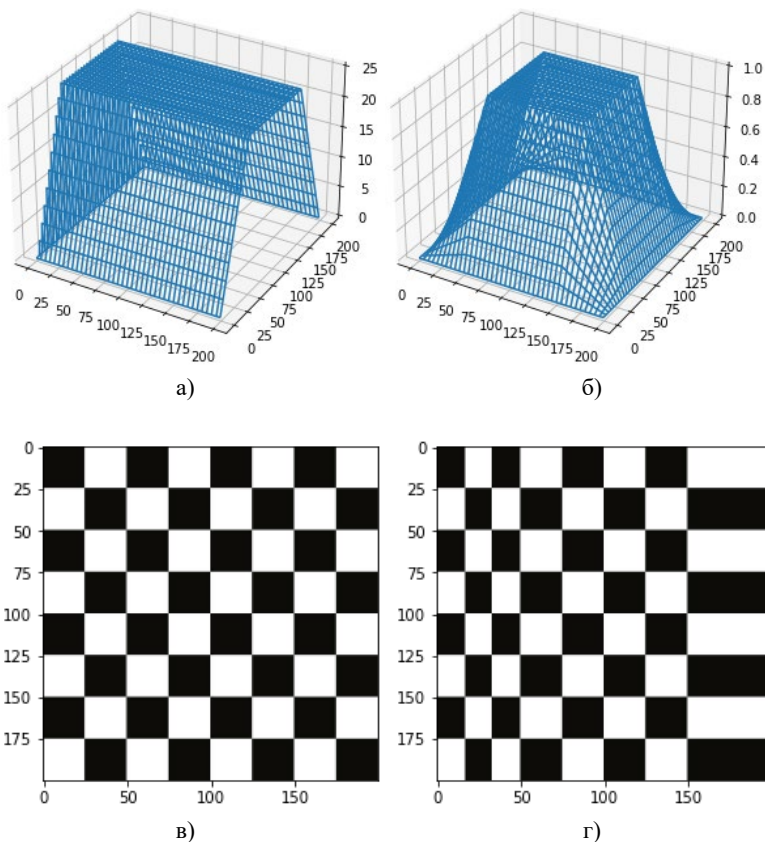


Рис. 2. Вид функции деформации, заданной: а) на основе маски  $h_1$  только вдоль одной из осей координат; б) масок  $h_1$  и  $h_2$  в полной мере; в) эталонное изображение; г) результат его преобразования при помощи функции ДП (а)

Собственно, функция деформирующих искажений может задаваться в целочисленном виде как квазидетерминированная функция со случайными параметрами и наложенной на нее маской (6).

При этом рассматривались два варианта. Первый вариант предполагает использование функции вида:

$$v_{1,2}(\mathbf{z}) = S_{a1,2} h(\mathbf{z}) \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{m}_{rnd})^T C_{rnd}^{-1}(\mathbf{z} - \mathbf{m}_{rnd})\right],$$

$$S_{a1,2} \in \{1; -1\}, \mathbf{m}_{rnd} \in \{\mathbf{m}_{T,L}, \mathbf{m}_{T,R}, \mathbf{m}_{B,L}, \mathbf{m}_{B,R}, \mathbf{m}_{C,C}\}, \quad (7)$$

$$C_{rnd} = \frac{NM}{3} \begin{pmatrix} 1 & \rho_{rnd} \\ \rho_{rnd} & 1 \end{pmatrix}, \quad -1 < \rho_{rnd} < 1,$$

где  $\mathbf{z} = (z_1, z_2)^T = (i, k)^T$  – вектор дискретных координат точки на изображении;  $S_{a1,2}$  – случайная величина, определяющая знак импульса ДИ;  $\mathbf{m}_{rnd}$  – случайным образом выбираемый центр размещения максимума функции ДИ по четырем углам сетки  $\Psi$  с некоторым отступом от края, а также в ее центре, определяемом вектором  $\mathbf{m}_{C,C} = (m_{C,x}, m_{C,y})^T$ ,  $m_{C,x} = \text{round}(N/2)$ ,  $m_{C,y} = \text{round}(M/2)$ ;  $C_{rnd}$  – положительно определенная матрица со случайным равновероятным в указанном диапазоне значений коэффициентом  $\rho_{rnd}$  и параметром влияния  $d_s$ , определяющем фактически дисперсию импульса ДИ. Фактически данная функция определяет форму импульса ДИ, знак которой задает различные направления выполняемой деформации.

Второй вариант предполагает использование функции вида:

$$v_{1,2}(\mathbf{z}) = S_{a1,2} h(\mathbf{z}) \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{m}_{rnd})^T C_{rnd}^{-1}(\mathbf{z} - \mathbf{m}_{rnd})\right] a(\mathbf{z}), \quad (8)$$

$$a(\mathbf{z}) = (z_1 - m_{rnd,x} + 1)(z_2 - m_{rnd,y} + 1).$$

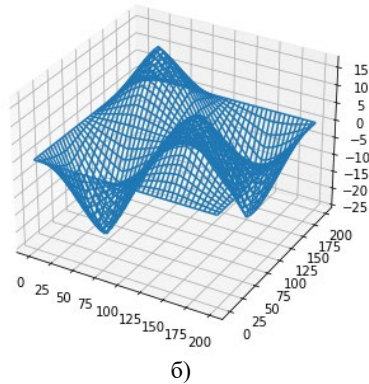
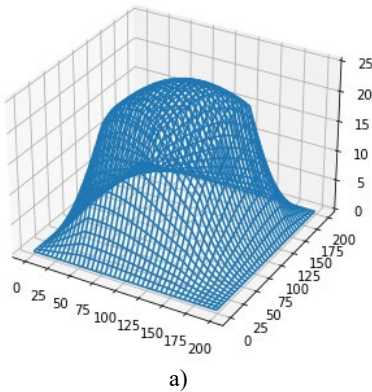
В отличие от предыдущего варианта здесь импульс ДИ обеспечивает максимальную по амплитуде деформацию в центрах квадрантов, размещающихся относительно центральной точки изображения.

После наложения маски к обеим функциям применялась нормализация с приведением максимального значения к величине амплитуды  $A_m$ :

$$\widehat{v}_{1,2}(\mathbf{z}) = A_m v_{1,2}(\mathbf{z}) / \max_{\Psi} v_{1,2}(\mathbf{z}).$$

На рисунке 3(а) представлен вид функции ДИ (7) с наложенной на нее стандартной маской (6) (рисунок 2(б)) и с размещением в центре  $\mathbf{m}_{rnd} = \mathbf{m}_{C,C}$ , а на рисунке 3в – соответствующее искаженное изображение изначально регулярной шахматной структуры (рисунок 2(в)). На рисунке 3(б) представлен вид функции ДИ (8) с наложенной маской (6) и с размещением в центре  $\mathbf{m}_{rnd} = \mathbf{m}_{C,C}$ , а на рисунке 3(г) – соответствующее искаженное изображение изначально регулярной шахматной структуры. Анализ представленных изображений показывает, что предложенные алгоритмы реализуют достаточно плавное их искажение. При этом все структурные элементы – сегменты исходного изображения сохраняются даже при весьма значительной амплитуде вносимых деформаций. Также и сохраняются границы между ними.

Следует отметить, что для повышения быстродействия процесса генерации ДИ при необходимости можно ограничиться использованием общей функции деформации по обеим координатам  $v_1(\mathbf{z}) = v_2(\mathbf{z}) = v(\mathbf{z})$ .



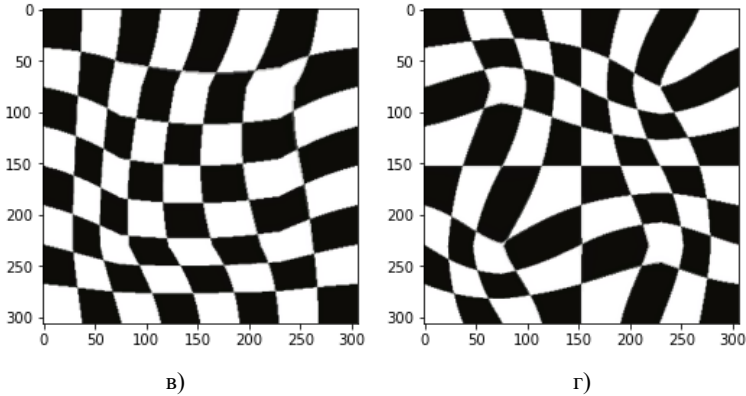


Рис. 3. Примеры задания функций деформации и соответствующим образом деформированных эталонных изображений

Как уже упоминалось выше, одной из целей данной работы было создание относительно быстрого алгоритма ДП, ориентированного на снижение затрат времени при ИРД в задачах классификации изображений. Основная идея предлагаемого алгоритма состоит в использовании плавных деформаций импульсного характера с реализацией оператора перестановки элементов исходного изображения в элементы трансформируемого изображения в рамках наложенной ограничивающей маски. Для оценки получаемого выигрыша по времени выполнения ДП было проведено сравнительное исследование этого алгоритма с алгоритмом ElasticTransform. При этом рассмотрено две известные программные реализации последнего, размещенная в виде исходного кода в открытом доступе на ресурсе Kaggle [32] и размещенная в составе библиотеки Albumentations [17]. Предлагаемый алгоритм также был реализован в двух вариантах. Первый вариант основан на генерации различных функций деформации  $v_1(\mathbf{z}), v_2(\mathbf{z})$  для каждой пространственной координаты в соответствии с (8). Второй вариант реализует использование общей функции деформации  $v_1(\mathbf{z}) = v_2(\mathbf{z}) = v(\mathbf{z})$  для каждой координаты. Следует отметить, что визуально получаемые от применения этих вариантов результаты отличаются не существенно.

Сравнение проводилось для разных размеров изображений и разных значений гиперпараметров алгоритмов, отвечающих за степень плавности и амплитуду деформации. Для предлагаемого алгоритма таким параметром является амплитуда ДП, определяемая относительно размера изображения  $N = N_x = N_y$  как  $A_m = k_s \times N$ , где

$k_s$  – варьируемый в экспериментах коэффициент  $0 < k_s < 1$ . Для алгоритма ElasticTransform варьируемым параметром являлась величина  $\sigma = k_s \times N$ , которая определяет размеры окна сглаживающих фильтров, используемых в этом алгоритме. При этом чем больше величина  $k_s$ , тем получаемые деформации являются более плавными и, напротив, при малых значениях этого коэффициента вносимые деформации становятся более изрезанными вплоть до потери структурного сходства элементов. В ходе вычислительного эксперимента замеры времени производились для эталонного изображения, вид которого представлен на рисунке 2, которое 1000 раз подвергалось ДП на стационарном ПК. Полученные результаты представлены в таблице 1 и носят, естественно, относительный характер, позволяя, тем не менее, проводить сопоставительный анализ алгоритмов.

Таблица 1. Сравнение временных характеристик алгоритмов ДП при выполнении 1000 запусков алгоритма

Используемый алгоритм	Время (с), $N = 200$ , $A_m = 0,125N$ , $\sigma = 0,05N$	Время (с), $N = 400$ , $A_m = 0,125N$ , $\sigma = 0,05N$	Время (с), $N = 600$ , $A_m = 0,125N$ , $\sigma = 0,05N$	Время (с), $N = 400$ , $A_m = 0,0625N$ , $\sigma = 0,025N$	Время (с), $N = 400$ , $A_m = 0,25N$ , $\sigma = 0,1N$
ElasticTransform (Kaggle)	37,47	245,52	753,30	164,54	415,56
ElasticTransform (Albumentations)	17,27	113,29	364,87	71,84	206,61
Предлагаемый алгоритм (вариант 1)	6,00	34,46	83,26	34,33	34,61
Предлагаемый алгоритм (вариант 2)	3,58	19,99	48,41	19,93	19,76

Анализ показывает, что предлагаемый алгоритм позволяет получить повышение быстродействия от 2,9 до 4,4 раз для первого варианта реализации и от 4,8 до 7,6 для второго варианта в зависимости от размера изображения при фиксированных значениях гиперпараметров (первые три колонки). Также следует отметить практическую инвариантность времени его выполнения относительно амплитуды деформации, тогда как время выполнения алгоритма ElasticTransform существенно зависит от гиперпараметра  $\sigma$ ,

отвечающего за уровень вносимых деформаций (колонки 2, 4, 5 таблицы 1).

**3. Методика и результаты обучения нейронных сетей с использованием предлагаемого метода аугментации.** В ходе проведенных исследований решалась задача оценки влияния предлагаемого способа аугментации изображений в задаче классификации на основе обучения с нуля относительно легковесных глубоких нейронных сетей стандартной архитектуры. Таких сетей было рассмотрено две: сверточная сеть стандартной архитектуры (convolutional neural network, CNN) и сеть с остаточными связями (deep residual network, DRN). В зависимости от объема обучающих данных количественные характеристики слоев и используемые при обучении гиперпараметры подбирались различным образом.

Архитектура первой сети в виде набора повторяющихся фрагментов показана на рисунке 4(а). Здесь использованы следующие обозначения:  $N$  – размер входного изображения;  $k_n$  – размер ядра свертки;  $filters$  – количество фильтров (каналов) первого сверточного слоя, определяющее количество реализуемых карт признаков;  $N_1, \dots, N_5$  – умножающий коэффициент для количества фильтров на последующих слоях;  $S$  – величина сдвига при проведении свертки или пуллинга;  $dr$  – величина коэффициента слоя дропаута, вставленного после выполнения предшествующего слоя, отделенного чертой;  $R_n$  – количество нейронов в первом плотном слое;  $Num\_classes$  – количество классов изображений;  $Relu, Softmax$  – стандартные обозначения активационных функций, используемых в слоях нейронной сети;  $Ind\_classes$  – формируемый на выходе индекс класса. Во всех сверточных слоях использовалась стандартная batch-нормализация.

Архитектура второй сети показана на рисунке 4(б). Здесь использованы следующие обозначения:  $N$  – размер входного изображения;  $k_n$  – размер ядра свертки;  $filters$  – количество фильтров (каналов) первого сверточного слоя, определяющее количество реализуемых карт признаков;  $N_1, \dots, N_5$  – умножающий коэффициент для количества фильтров на последующих слоях;  $S$  – величина сдвига при проведении свертки или пуллинга;  $dr$  – величина коэффициента слоя дропаута, вставленного после выполнения предшествующего слоя, отделенного чертой;  $R_n$  – количество нейронов в первом плотном слое;  $Num\_classes$  – количество классов изображений;  $Relu, Softmax$  – стандартные

обозначения активационных функций, используемых в слоях нейронной сети; Ind\_classes – формируемый на выходе индекс класса.



а) б)  
Рис. 4. Архитектура используемых нейронных сетей



Методика подготовки данных для обучения нейронных сетей состояла в следующем. Использовались три исходных обучающих выборки изображений, приведенных к общему размеру по обем осям  $w_{size}$ :

- малая выборка  $B_1$  изображений 3 классов животных, содержащая по 1000 экземпляров изображений в каждом классе (всего 3000);
- средняя по объему выборка  $B_2$  изображений 3 классов животных, содержащая по 2000 экземпляров изображений в каждом классе (всего 6000);
- средняя по объему выборка  $B_3$  изображений 10 классов животных, содержащая по 2000 экземпляров изображений в каждом классе (всего 20000).

Выборка  $B_1$  (как наиболее важная с точки зрения исследования) была сформирована в двух вариантах с различающимися изображениями одних и тех же классов, полученными из разных источников.

Каждая выборка  $B_1$ ,  $B_2$ ,  $B_3$  подвергалась преобразованиям, направленным на искусственное размножение обучающих данных. Сначала на основе исходной формировалась выборка увеличенного вдвое объема путем выполнения зеркального отражения каждого изображения. Таким образом, были получены выборки:  $B_1^{flip}$  – 6000 изображений;  $B_2^{flip}$  – 12000 изображений;  $B_3^{flip}$  – 40000 изображений. Затем уже эти выборки подвергались аугментации на основе изложенного выше алгоритма внесения деформирующих искажений, что позволило, в свою очередь, получить выборки  $B_1^{flip+aug}$  – 12000 изображений;  $B_2^{flip+aug}$  – 24000 изображений;  $B_3^{flip+aug}$  – 80000 изображений. Выборки  $B_1$ ,  $B_2$  и преобразованные от них выборки разбивались на обучающую и валидационную, исходя из соотношения 80% и 20%. Выборка  $B_3$  и преобразованные от нее выборки разбивались на обучающую и валидационную, исходя из соотношения 90% и 10%. Для тестирования всегда использовалась подвыборка, содержащая, соответственно, 20% из  $B_1$ ,  $B_2$  и 10% из  $B_3$ , т.е. исключительно состоящая из оригинальных, не преобразованных каким-либо образом изображений. Поэтому при обучении по исходным, не преобразованным выборкам,  $B_2$ ,  $B_3$  тестирующие и валидационные выборки совпали. Для выборок  $B_1$ ,  $B_2$  и производных

от них выборки использовались два значения размера изображения  $w_{size} = 128$  и  $w_{size} = 200$ . Для выборки  $B_3$  и ее аугментированных вариантов использовалось значение  $w_{size} = 128$  ввиду значительного роста времени обучения.

Аугментация ДИ осуществлялась в варианте (7) с параметром  $A_m = 16$  для  $w_{size} = 128$  и  $A_m = 25$  для  $w_{size} = 200$ .

Для первой сети при обучении по малым выборкам  $B_1$ ,  $B_2$  и их производных задавались следующие гиперпараметры. Значение исходного количества `filters` устанавливалось равным 32. Использовались пять вычислительных сверточных блоков (рисунок 4а) с коэффициентами умножения числа фильтров  $N1=1, N2=N3=2, N4=N5=4$ . Размер ядра фильтров свертки устанавливался равным  $kn=3$  при  $w_{size} = 128$  и  $kn=5$  при  $w_{size} = 200$ . В слоях сетей использовалась L2-регуляризация с коэффициентом  $l2_{val}=0,01$ , а также регуляризация на основе метода стохастического исключения весовых связей (метод `dropout`) со значением вероятности исключения  $dr=0,2$ . При обучении использовался оптимизатор RMSprop с начальной скоростью  $lr_{val}=0,0004$ . Процесс обучения занимал 200 эпох и сопровождался постепенным понижением скорости обучения до нижнего предела 0,00001. При обучении по выборке  $B_3$  и производным от нее выборкам задавались аналогичные параметры, с тем только отличием, что коэффициенты умножения числа фильтров задавались как  $N1=1, N2=2, N3=4, N4=8, N5=16$ . В слоях сетей использовалась L2-регуляризация с коэффициентом  $l2_{val}=0,001$ , при этом регуляризация на основе метода стохастического исключения весовых связей не проводилась ( $dr=0,0$ ). Процесс обучения занимал 100 эпох и сопровождался постепенным понижением скорости обучения до нижнего предела 0,00001.

Для второй сети аналогично при обучении по малым выборкам  $B_1$ ,  $B_2$  и их производных задавались следующие гиперпараметры. Значение исходного количества `filters` устанавливалось равным 32. Использовались пять вычислительных сверточных блоков (рисунок 4(б)) с коэффициентами умножения числа фильтров  $N1=1, N2=N3=2, N4=N5=4$ , размер ядра фильтров свертки устанавливался равным  $kn=3$  при  $w_{size} = 128$  и при  $w_{size} = 200$ . В слоях сетей аналогично использовалась L2-регуляризация с коэффициентом  $l2_{val}=0,01$ , а также регуляризация на основе метода

стохастического исключения весовых связей (метод dropout) со значением вероятности исключения  $dr=0,2$ . При обучении использовался оптимизатор RMSprop с начальной скоростью  $lr_{val}=0,0004$ . Процесс обучения занимал 200 эпох и сопровождался постепенным понижением скорости обучения до нижнего предела 0,00001.

При обучении по выборке  $B_3$  и производным от нее выборкам задавались аналогичные параметры, с тем только отличием, что коэффициенты умножения числа фильтров задавались как  $N1=1, N2=2, N3=4, N4=8, N5=16$ . В слоях сетей аналогично использовалась L2-регуляризация с коэффициентом  $l2_{val}=0,001$ , при этом регуляризация на основе метода стохастического исключения весовых связей не проводилась ( $dr=0,0$ ). Процесс обучения занимал 100 эпох и сопровождался постепенным понижением скорости обучения до нижнего предела 0,00001.

Как уже отмечалось, в процессе обучения был реализован план ступенчатого изменения скорости обучения в указанных пределах с сохранением весовых коэффициентов сети, которая показала лучшие по отношению к прошлым результаты по валидационной подвыборке. Последняя из сохраненных сетей использовалась для тестирования. Несмотря на очевидные признаки переобучения сетей на завершающей части обучения, определенное улучшение качества классификации происходило практически до последней эпохи.

В таблицах 2 – 4 представлены полученные в ходе двухэтапной аугментации результаты точности классификации (метрика assigasy) для сетей классов CNN и DRN представленной архитектуры.

Таблица 2. Результаты тестирования при обучении по выборке  $B_1$  и производным от нее выборкам (3 класса,  $l2=0,01, dr=0,2$ , в ячейках со значениями: сверху – первый вариант  $B_1$ , внизу – второй вариант  $B_1$ )

Размер изображения и параметры ядер свертки	Обучение по выборке $B_1$		Обучение по выборке $B_1^{flip}$		Обучение по выборке $B_1^{flip+aug}$	
	CNN	DRN	CNN	DRN	CNN	DRN
$w_{size} = 128, kn=3$	85,57 86,50	85,83 83,66	88,67 88,83	90,50 87,00	<b>90,68</b> <b>89,60</b>	<b>91,00</b> <b>89,67</b>
$w_{size} = 200, kn=5$	88,00 87,83	88,17 00,00	90,16 89,00	90,17 87,17	<b>92,60</b> <b>89,83</b>	<b>92,67</b> <b>87,83</b>

Таблица 3. Результаты тестирования при обучении по выборке  $B_2$  и производным от нее выборкам (3 класса,  $l_2=0,01$ ,  $dr=0,2$ )

Размер изображения и параметры ядер свертки	Обучение по выборке $B_2$		Обучение по выборке $B_2^{flip}$		Обучение по выборке $B_2^{flip+aug}$	
	CNN	DRN	CNN	DRN	CNN	DRN
$w_{size} = 128, kn=3$	93,75	94,17	94,42	95,50	<b>95,92</b>	<b>95,58</b>
$w_{size} = 200, kn=5$	93,58	93,33	95,08	94,83	<b>95,83</b>	<b>95,50</b>

Таблица 4. Результаты тестирования при обучении по выборке  $B_3$  и производным от нее выборкам (10 классов,  $l_2=0,001$ ,  $dr=0,0$ )

Размер изображения и параметры ядер свертки	Обучение по выборке $B_3$		Обучение по выборке $B_3^{flip}$		Обучение по выборке $B_3^{flip+aug}$	
	CNN	DRN	CNN	DRN	CNN	DRN
$w_{size} = 128, kn=3$	89,25	84,85	91,28	89,90	<b>92,05</b>	<b>92,60</b>

Анализ представленных в таблицах 2–4 результатов показывает, что во всех проведенных экспериментах аугментация, выполненная только на основе стандартной процедуры зеркального отражения, дает прирост в точности (частоты правильного распознавания). Этот прирост особенно заметен при размножении малой выборки  $B_1$  и, в отдельных случаях, достигает уровня 3,0%. Следующий этап ИРД, выполненный по отношению как к исходным, так и к зеркально отраженным выборкам, также показал наличие прироста точности классификации во всех экспериментах, который аналогично особенно заметен для исходной выборки  $B_1$ . Здесь его значение достигает значений до 2,5%. Для выборки  $B_2$  прирост, естественно, менее значителен, особенно для сети DRN. Для выборки  $B_3$ , напротив, наиболее значимое увеличение точности 2,7% получено для сети DRN. В целом можно констатировать, что суммарный прирост точности классификации по результатам выполнения обоих этапов аугментации имеет значения, в основномходящие до 5%. Это, на наш взгляд является вполне удовлетворительным результатом для фиксируемого диапазона значений точности классификации.

Анализ уровня статистической значимости полученных результатов проводился по стандартным соотношениям для оцененных вероятностей и представлен в таблице 5 в виде максимального значения модуля отклонения оцененной точности классификации от возможной истинной. Объем независимой тестирующей подвыборки исходных (не преобразованных) изображений составлял, соответственно, 20% из  $B_1$ ,  $B_2$  и 10% из  $B_3$ .

Таблица 5. Результаты анализа статистической значимости

Объем тестирующей подвыборки	Обучение и тестирование по выборке $B_1$ (600 примеров)		Обучение и тестирование по выборке $B_2$ (1200 примеров)		Обучение и тестирование по выборке $B_3$ (2000 примеров)	
	10 %	5 %	10 %	5 %	10 %	5 %
Максимальное значение модуля отклонения от истинного значения	1,46 %	1,74 %	1,03 %	1,23 %	0,80 %	0,96 %

Анализ представленных результатов показывает, что полученный для каждого эксперимента суммарный по результатам выполнения двух этапов прирост точности классификации является статистически значимым. Также, с учетом смещения всех полученных оценок в большую сторону, статистически значимым является и прирост, полученный только на основе аугментации с внесением деформирующих искажений.

Следует также отметить, что различия в эффективности применения сетей различной архитектуры в проведенных экспериментах не являются принципиальными с учетом того, что поставленная задача состоит не в нахождении лучшей сети, а в демонстрации возможностей предлагаемого метода аугментации данных.

**4. Заключение.** Таким образом, в работе предложен подход к обучению глубоких нейронных сетей для классификации изображений в условиях малой и средней по объему выборки, основанный на применении техники аугментации обучающих данных с использованием зеркальных отражений и деформирующих преобразований обрабатываемых изображений. Предложена математическая модель и обоснован реализующий ее алгоритм

внесения деформирующих искажений, основанные на выполнении циклической перестановки элементов исходного изображения в соответствие с используемой функцией пространственной деформации. Показано, что его применение при любых уровнях деформации сохраняет структурную основу исходного изображения, что делает его пригодным для проведения аугментации. Показано также, что быстроедействие алгоритма существенно выше по сравнению с известными реализациями аналогичного характера. В ходе всех экспериментов с обучением нейронных сетей различной архитектуры установлено наличие прироста точности классификации при применении процедуры стандартной аугментации на основе зеркального отражения исходных изображений обучающей выборки с последующим применением аугментации на основе предложенного алгоритма деформирующего преобразования. Указанный прирост для рассмотренных архитектур нейронных сетей наиболее заметен при обучении по относительно малой выборке (порядка  $10^3$  примеров на каждый класс).

Дальнейшие исследования, на наш взгляд, целесообразно сосредоточить на исследовании возможностей предлагаемого подхода для повышения качества обучения в задаче семантической сегментации, а также его сравнении с известными алгоритмами, которые ранее использовались при решении данной задачи.

### Литература

1. Chawla N.V., Lazarevic A., Hall L.O., Bowyer K.W. SMOTEBoost: Improving Prediction of the Minority Class in Boosting // 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD). 2003. pp. 107–119. DOI: 10.1007/978-3-540-39804-2\_12.
2. Minaee S., Luo P., Lin Zh., Bowyer K. Going deeper into face detection: A survey // arXiv preprint. 2021. DOI: 10.48550/arXiv.2103.14983.
3. Ciresan D.C., Meier U., Gambardella L.M., Schmidhuber J. Deep, Big, Simple Neural Nets For Handwritten Digit Recognition // Neural computation. 2010. vol. 22. no. 12. pp. 3207–3220. DOI: 10.1162/NECO\_a\_00052.
4. Tao X., Zhang D., Ma W., Liu X., Xu D. Automatic Metallic Surface Defect Detection and Recognition with Convolutional Neural Networks // Applied Sciences. 2018. vol. 8. no. 9. pp. 1575–1590. DOI: 10.3390/app8091575.
5. Shorten C., Khoshgoftaar T.M. Survey on Image Data Augmentation for Deep Learning // Journal of Big Data. 2019. vol. 6. no. 1. pp. 1–48. DOI: 10.1186/s40537-019-0197-0.
6. Wang W., Xie E., Li X., Fan, D. P., Song, K., Liang, D., Lu T., Luo P., Shao, L. Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction Without Convolutions // Proceedings of the IEEE/CVF international conference on computer vision. 2021. pp. 568–578. DOI: 10.1109/ICCV48922.2021.00061.
7. Качалин С.В. Повышение устойчивости обучения больших нейронных сетей дополнением малых обучающих выборок примеров-родителей, синтезированными биометрическими примерами-потомками // Труды научно-

- технической конференции кластера пензенских предприятий, обеспечивающих безопасность информационных технологий. 2014. Т. 9. С. 32–35.
8. Акимов А.В., Сирота А.А. Модели и алгоритмы искусственного размножения данных для обучения алгоритмов распознавания лиц методом Виолы–Джонса // Компьютерная оптика. 2016. Т. 40. № 6. С. 911–918. DOI: 10.18287/2412-6179-2016-40-6-911-918.
  9. Небаба С.Г., Захарова А.А. Алгоритм построения деформируемых 3D моделей лица и обоснование его применимости в системах распознавания личности. Труды СПИИРАН. 2017. Т. 52. С. 157–179. DOI: 10.15622/sp.52.8.
  10. Сирота А.А., Донских А.О., Акимов А.В., Минаков Д.А. Смешанные ядерные оценки многомерных распределений и их применение в задачах машинного обучения для классификации биологических объектов на основе спектральных измерений // Компьютерная оптика. 2019. Т. 43. № 4. С. 677–691. DOI: 10.18287/2412-6179-2019-43-4-677-691.
  11. Дагаева М.В., Сулейманов М.А., Катаева Д.В., Катаев, А.С., Кирпичников А.П. Технология построения отказоустойчивых нейросетевых моделей распознавания рукописных символов в системах биометрической аутентификации // Вестник Технологического университета. 2018. Т. 21. № 2. С. 133–138.
  12. Емельянов С.О., Иванова А.А., Швец Е.А., Николаев Д.П. Методы аугментации обучающих выборок в задачах классификации изображений // Сенсорные системы. 2018. Т. 32. № 3. С. 236–245. DOI: 10.1134/S0235009218030058.
  13. Рюмина Е.В., Рюмин Д.А., Маркитантов М.В., Карлов А.А. Метод генерации обучающих данных для компьютерной системы обнаружения защитных масок на лицах людей // Компьютерная оптика. 2022. Т. 46. № 4. С. 603–611. DOI: 10.18287/2412-6179-CO-1039.
  14. Камалова Ю.Б., Андриянов Н.А. Распознавание микроскопических изображений пыльцевых зерен с помощью сверточной нейронной сети VGG-16 // Вестник Южно-Уральского государственного университета. Серия: Компьютерные технологии, управление, радиоэлектроника. 2022. Т. 22. № 3. С. 39–46. DOI: 10.14529/ctcr220304.
  15. Ковун В.А., Каширина И.Л. Использование нейронной сети W-Net в металлографическом анализе образцов стали // Вестник ВГУ (Системный анализ и информационные технологии). 2022. № 1. С. 101–110. DOI: 10.17308/sait.2022.1/9205.
  16. Simard P.Y., Steinkraus D., Platt J.C. Best practices for convolutional neural networks applied to visual document analysis // In Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR '03). 2003. vol. 2. pp. 1–6.
  17. Buslaev A., Igllovikov V.I., Khvedchenya E., Parinov A., Druzhinin M., Kalinin A.A. Albumentations: Fast and flexible image augmentations. Information. 2020. vol. 11. no. 2. pp. 1–20. DOI: 10.3390/info11020125.
  18. Hasan S.M.K., Linte C.A. U-NetPlus: A modified encoder-decoder U-Net architecture for semantic and instance segmentation of surgical instruments from laparoscopic images // 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2019. pp. 7205–7211.
  19. Keysers D., Deselaers T., Gollan C., Ney H. Deformation models for image recognition // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007. vol. 29(8). pp. 1422–1435. DOI: 10.1109/TPAMI.2007.1153.
  20. Felzenszwalb P., McAllester D., Ramanan D. A discriminatively trained, multiscale, deformable part model // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2008. pp. 1–8. DOI: 10.1109/CVPR.2008.4587597.

21. Wiskott L., Fellous J.-M., Kruger N., von der Malsburg C. Face Recognition by Elastic Bunch Graph Matching // Proceedings of International Conference on Image Processing. 1997. vol. 1. pp. 129–132. DOI: 10.1109/ICIP.1997.647401.
22. Li X., Li W., Ren D., Zhang H., Wang M., Zuo W. Enhanced Blind Face Restoration with Multi-Exemplar Images and Adaptive Spatial Feature Fusion // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. pp. 2706–2715. DOI: 10.1109/CVPR42600.2020.00278.
23. Deng Y., Yang J., Tong X. Deformed Implicit Field: Modeling 3D Shapes With Learned Dense Correspondence // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. pp. 10286–10296. DOI: 10.48550/arXiv.2011.13650.
24. Venkatesh S., Ramachandra R., Raja K., Busch Ch. Face Morphing Attack Generation and Detection: A Comprehensive Survey // IEEE Transactions on Technology and Society. 2021. vol. 2. no. 3. pp. 128–145. DOI: 10.1109/TTS.2021.3066254.
25. Scherhag U., Rathgeb C., Merkle J., Busch C. Deep Face Representations for Differential Morphing Attack Detection // IEEE Transactions on Information Forensics and Security. 2020. vol. 15. pp. 3625–3639. DOI: 10.1109/TIFS.2020.2994750.
26. Ling H., Kreis K., Li D., Kim S.W., Torralba A., Fidler S. EditGAN: High-Precision Semantic Image Editing // Advances in Neural Information Processing Systems. 2021. vol. 34. pp. 16331–16345. DOI: 10.48550/arXiv.2111.03186.
27. Wang S.Y., Bau D., Zhu J.Y. Rewriting Geometric Rules of a GAN // ACM Transactions on Graphics (TOG). 2022. vol. 41. no. 4. pp. 1–16. DOI: 10.48550/arXiv.2207.14288.
28. Акимов А.В., Дрюченко М.А., Сирота А.А. Модели и алгоритмы внесения деформирующих искажений на изображениях с использованием радиально-базисных функций // Вестник ВГУ (Системный анализ и информационные технологии). 2014. № 1. С. 130–137.
29. Захарова А.А., Небаба С.Г., Завьялов Д.А. Алгоритмическое и программное обеспечение для повышения эффективности обработки многомерных гетерогенных данных // Программирование. 2019. № 4. С. 64–70. DOI: 10.1134/S0132347419040101.
30. Buckley M.J. Fast computation of a discretized thin-plate smoothing spline for image data // Biometrika. 1994. vol. 81. no. 2. pp. 247–258. DOI: 10.2307/2336955.
31. Sastry S.P., Zala V., Kirby R.M. Thin-plate-spline curvilinear meshing on a calculus-of-variations framework // Procedia Engineering. 2015. vol. 124. pp. 135–147. DOI: 10.1016/j.proeng.2015.10.128.
32. Elastic Transform for Data Augmentation. URL: <https://www.kaggle.com/code/bguberfain/elastic-transform-for-data-augmentation> (accessed: 30.10.2023).

**Сирота Александр Анатольевич** — д-р техн. наук, профессор, заведующий кафедрой, кафедра технологий обработки и защиты информации факультета компьютерных наук, ФГБОУ ВО «Воронежский государственный университет». Область научных интересов: синтез и анализ систем сбора и обработки информации, методы и технологии компьютерного моделирования информационных процессов и систем, машинное обучение, компьютерная обработка изображений, нейронные сети и нейросетевые технологии в системах принятия решений. Число научных публикаций — 303. [sir@cs.vsu.ru](mailto:sir@cs.vsu.ru); Университетская площадь, 1, 394018, Воронеж, Россия; р.т.: +7(903)030-6943.



**Акимов Алексей Викторович** — канд. физ.-мат. наук, старший преподаватель, кафедра технологий обработки и защиты информации факультета компьютерных наук, ФГБОУ ВО «Воронежский государственный университет». Область научных интересов: распознавание изображений, машинное обучение. Число научных публикаций — 28. akimov@vsu.ru; Университетская площадь, 1, 394018, Воронеж, Россия; р.т.: +7(903)030-6943.

**Отырба Ростислав Русланович** — аспирант, кафедра технологий обработки и защиты информации факультета компьютерных наук, ФГБОУ ВО «Воронежский государственный университет». Область научных интересов: машинное обучение, глубокое обучение, компьютерное зрение, обработка естественного языка. Число научных публикаций — 6. otyrba@cs.vsu.ru; Университетская площадь, 1, 394018, Воронеж, Россия; р.т.: +7(903)854-9545.

A. SIROTA, A. AKIMOV, R. OTYRBA  
**IMAGE WARPING AND ITS APPLICATION FOR DATA  
AUGMENTATION WHEN TRAINING DEEP NEURAL  
NETWORKS**

*Sirota A., Akimov A., Otyrba R. Image Warping and Its Application for Data Augmentation when Training Deep Neural Networks.*

**Abstract.** The paper focuses on the improvement of the quality of learning for deep neural networks for a small data set in a classification task. One of the possible approaches to improve the quality of learning is researched which is based on the use of data augmentation (artificial reproduction of the data set) by image warping. The presented mathematical model and fast algorithm for warping make it possible to transform the original image while preserving its structural basis. The proposed algorithm is used to augment image data sets containing a small number of training samples. The augmentation consists of two stages including horizontal mirroring and warping of each of the samples. The effectiveness of such augmentation is tested through the training of neural networks of various types: convolutional neural networks (CNN) of a standard architecture and deep residual networks (DRN). A specific feature of the implemented approach for the solution of the problem under consideration consists in the refusal to use pre-trained neural networks with a large number of layers as well as further transfer learning, since their application incurs costs in terms of the computational resources. The paper shows that the efficiency of image classification when implementing the proposed method of augmenting training data on small and medium-sized data sets increases to statistically significant values of the metric used.

**Keywords:** deep neural networks, training data augmentation, image warping, efficiency of deep neural networks.

## References

1. Chawla N.V., Lazarevic A., Hall L.O., Bowyer K.W. SMOTEBoost: Improving Prediction of the Minority Class in Boosting. 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD). 2003. pp. 107–119. DOI: 10.1007/978-3-540-39804-2\_12.
2. Minaee S., Luo P., Lin Zh., Bowyer K. Going deeper into face detection: A survey. arXiv preprint. 2021. DOI: 10.48550/arXiv.2103.14983.
3. Ciresan D.C., Meier U., Gambardella L.M., Schmidhuber J. Deep, Big, Simple Neural Nets For Handwritten Digit Recognition. Neural computation. 2010. vol. 22. no. 12. pp. 3207–3220. DOI: 10.1162/NECO\_a\_00052.
4. Tao X., Zhang D., Ma W., Liu X., Xu D. Automatic Metallic Surface Defect Detection and Recognition with Convolutional Neural Networks. Applied Sciences. 2018. vol. 8. no. 9. pp. 1575–1590. DOI: 10.3390/app8091575.
5. Shorten C., Khoshgoftaar T.M. Survey on Image Data Augmentation for Deep Learning. Journal of Big Data. 2019. vol. 6. no. 1. pp. 1–48. DOI: 10.1186/s40537-019-0197-0.
6. Wang W., Xie E., Li X., Fan D.P., Song K., Liang D., Lu T., Luo P., Shao L. Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction Without Convolutions. Proceedings of the IEEE/CVF international conference on computer vision. 2021. pp. 568–578. DOI: 10.1109/ICCV48922.2021.00061.

7. Kachalin S.V. [Improving the stability of large neural net-works by extending small training sets of parent samples with synthesized biometric descendant samples]. Trudy nauchno-tehnicheskoy konferencii klastera penzenskih predpriyatij, obespechivajushhij bezopasnost' informacionnyh tehnologij – Proceedings of the Scientific and Technical Conference of The cluster of Penza Enterprises Providing Security of Information Technologies. 2014. vol. 9. pp. 32–35. (In Russ.).
8. Akimov, A.V., Sirota A.A. [Synthetic data generation models and algorithms for training image recognition algorithms using the Viola-Jones framework]. Komp'juternaja optika – Computer Optics. 2016. vol. 40. no. 6. pp. 911–918. DOI: 10.18287/2412-6179-2016-40-6-911-918. (In Russ.).
9. Nebaba S.G., Zakharova A.A. [An Algorithm for Building Deformable 3d Human Face Models and Justification of its Applicability for Recognition Systems]. Trudy SPIIRAN – SPIIRAS Proceedings. 2017. vol. 52. pp. 157–179. DOI: 10.15622/sp.52.8. (In Russ.).
10. Sirota A.A., Donskikh A.O., Akimov A.V., Minakov D.A. [Multivariate mixed kernel density estimators and their application in machine learning for classification of biological objects based on spectral measurements]. Komp'juternaja optika – Computer Optics. 2019. vol. 43. no. 4. pp. 677–691. DOI: 10.18287/2412-6179-2019-43-4-677-691. (In Russ.).
11. Dagaeva M.V., Sulejmanov M.A., Kataseva D.V., Katasyov, A.S., Kirpichnikov A.P. [Technology for building fault-tolerant neural network models for recognizing handwritten characters in biometric authentication systems]. Vestnik Tehnologicheskogo universiteta – Bulletin of the Technological University. 2018. vol. 21. no. 2. pp. 133–138. (In Russ.).
12. Emel'janov S.O., Ivanova A.A., Shvec E.A., Nikolaev D.P. [Methods of augmentation of training samples in image classification problems]. Sensornye sistemy – Sensory Systems. 2018. vol. 32. no. 3. pp. 236–245. DOI: 10.1134/S0235009218030058. (In Russ.).
13. Rjumina E.V., Rjumin D.A., Markitantov M.V., Karpov A.A. [A method for generating training data for a computer system for detecting protective masks on people's faces]. Komp'juternaja optika – Computer Optics 2022. vol. 46. no. 4. pp. 603–611. DOI: 10.18287/2412-6179-CO-1039. (In Russ.).
14. Kamalova Ju.B., Andrijanov N.A. [Recognition of microscopic images of pollen grains using the convolutional neural network VGG-16]. Vestnik Juzhno-Ural'skogo gosudarstvennogo universiteta. Serija: Komp'juternye tehnologii, upravlenie, radioelektronika – Bulletin of South Ural State University, Series: Computer Technologies, Automatic Control and Radioelectronics. 2022. vol. 22. no. 3. pp. 39–46. DOI: 10.14529/ctcr220304. (In Russ.).
15. Kovun V.A., Kashirina I.L. [Using the W-Net neural network in metallographic analysis of steel samples]. Vestnik VGU (Sistemnyj analiz i informacionnye tehnologii) – Vestnik VSU (System Analysis and Information Technology). 2022. no. 1. pp. 101–110. DOI: 10.17308/sait.2022.1/9205. (In Russ.).
16. Simard P.Y., Steinkraus D., Platt J.C. Best practices for convolutional neural networks applied to visual document analysis. In Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR '03). 2003. vol. 2. pp. 1–6.
17. Buslaev A., Igllovikov V.I., Khvedchenya E., Parinov A., Druzhinin M., Kalinin A.A. Albumentations: Fast and flexible image augmentations. Information. 2020. vol. 11. no. 2. pp. 1–20. DOI: 10.3390/info11020125.
18. Hasan S.M.K., Linte C.A. U-NetPlus: A modified encoder-decoder U-Net architecture for semantic and instance segmentation of surgical instruments from

- laparoscopic images. 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2019. pp. 7205–7211.
19. Keysers D., Deselaers T., Gollan C., Ney H. Deformation models for image recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2007. vol. 29(8). pp. 1422–1435. DOI: 10.1109/TPAMI.2007.1153.
  20. Felzenszwalb P., McAllester D., Ramanan D. A discriminatively trained, multiscale, deformable part model. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2008. pp. 1–8. DOI: 10.1109/CVPR.2008.4587597.
  21. Wiskott L., Fellous J.-M., Kruger N., von der Malsburg C. Face Recognition by Elastic Bunch Graph Matching. *Proceedings of International Conference on Image Processing*. 1997. vol. 1. pp. 129–132. DOI: 10.1109/ICIP.1997.647401.
  22. Li X., Li W., Ren D., Zhang H., Wang M., Zuo W. Enhanced Blind Face Restoration with Multi-Exemplar Images and Adaptive Spatial Feature Fusion. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020. pp. 2706–2715. DOI: 10.1109/CVPR42600.2020.00278.
  23. Deng Y., Yang J., Tong X. Deformed Implicit Field: Modeling 3D Shapes With Learned Dense Correspondence. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021. pp. 10286–10296. DOI: 10.48550/arXiv.2011.13650.
  24. Venkatesh S., Ramachandra R., Raja K., Busch Ch. Face Morphing Attack Generation and Detection: A Comprehensive Survey. *IEEE Transactions on Technology and Society*. 2021. vol. 2. no. 3. pp. 128–145. DOI: 10.1109/TTS.2021.3066254.
  25. Scherhag U., Rathgeb C., Merkle J., Busch C. Deep Face Representations for Differential Morphing Attack Detection. *IEEE Transactions on Information Forensics and Security*. 2020. vol. 15. pp. 3625–3639. DOI: 10.1109/TIFS.2020.2994750.
  26. Ling H., Kreis K., Li D., Kim S.W., Torralba A., Fidler S. EditGAN: High-Precision Semantic Image Editing. *Advances in Neural Information Processing Systems*. 2021. vol. 34. pp. 16331–16345. DOI: 10.48550/arXiv.2111.03186.
  27. Wang S.Y., Bau D., Zhu J.Y. Rewriting Geometric Rules of a GAN. *ACM Transactions on Graphics (TOG)*. 2022. vol. 41. no. 4. pp. 1–16. DOI: 10.48550/arXiv.2207.14288.
  28. Akimov AV, Dryuchenko MA, Sirota AA. [Models and algorithms for making distorting distortion in images using radial basis functions]. *Vestnik VGU (Sistemnyj analiz i informacionnye tehnologii) – Vestnik VSU (System Analysis and Information Technology)*. 2014. vol. 1. pp. 130–137. (In Russ.).
  29. Zaharova A.A., Nebaba S.G., Zav'jalov D.A. [Algorithmic and software to improve the efficiency of processing multidimensional heterogeneous data]. *Programmirovaniye – Programming*. 2019. no. 4. pp. 64–70. DOI: 10.1134/S0132347419040101. (In Russ.).
  30. Buckley M.J. Fast computation of a discretized thin-plate smoothing spline for image data. *Biometrika*. 1994. vol. 81. no. 2. pp. 247–258. DOI: 10.2307/2336955.
  31. Sastry S.P., Zala V., Kirby R.M. Thin-plate-spline curvilinear meshing on a calculus-of-variations framework. *Procedia Engineering*. 2015. vol. 124. pp. 135–147. DOI: 10.1016/j.proeng.2015.10.128.
  32. Elastic Transform for Data Augmentation. Available at: <https://www.kaggle.com/code/bguberfain/elastic-transform-for-data-augmentation> (accessed: 30.10.2023).

**Sirota Alexander** — Ph.D., Dr.Sci., Professor, Head of the department, Department of information security and processing technologies, faculty of computer sciences, Voronezh State University. Research interests: synthesis and analysis of systems for information

collection and processing, methods and technologies for computer modeling of information processes and systems, machine learning, computer image processing, neural networks and neural network technologies in decision-making systems. The number of publications — 303. sir@cs.vsu.ru; 1, Universitetskaya Sq., 394018, Voronezh, Russia; office phone: +7(903)030-6943.

**Akimov Aleksei** — Ph.D., Senior lecturer, Department of information security and processing technologies, faculty of computer sciences, Voronezh State University. Research interests: image recognition, machine learning. The number of publications — 28. akimov@vsu.ru; 1, Universitetskaya Sq., 394018, Voronezh, Russia; office phone: +7(903)030-6943.

**Otyrba Rostislav** — Postgraduate student, Department of information security and processing technologies, faculty of computer sciences, Voronezh State University. Research interests: machine learning, deep learning, computer vision, natural language processing. The number of publications — 6. otyrba@cs.vsu.ru; 1, Universitetskaya Sq., 394018, Voronezh, Russia; office phone: +7(903)854-9545.