

УДК 004.932

ВЫБОР ФУНКЦИИ АКТИВАЦИИ В СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЯХ ПРИ ПОВТОРНОЙ ИДЕНТИФИКАЦИИ ЛЮДЕЙ В СИСТЕМАХ ВИДЕОНАБЛЮДЕНИЯ

© 2022 г. Х. Чен^{a,b,*} (<http://orcid.org/0000-0003-4229-4505>),
С. Игнатьева^c (<http://orcid.org/0000-0002-9780-5731>),
Р. Богуш^{c,**} (<http://orcid.org/0000-0002-6609-5810>),
С. Абламейко^{d,e} (<http://orcid.org/0000-0001-9404-1206>)

^a Zhejiang Shuren University Shuren Str., 8, Hangzhou, China

^b International Science and Technology Cooperation Base of Zhejiang Province:
Remote Sensing Image Processing and Application
Hangzhou, 310000, China

^c Полоцкий государственный университет
ул. Блохина, д. 29, г. Новополоцк, 211446, Республика Беларусь

^d Белорусский государственный университет
n-т Независимости, д. 4, г. Минск, 220030, Республика Беларусь

^e Объединенный институт проблем информатики НАН Беларуси
ул. Сурганова, 6, г. Минск, 220012, Республика Беларусь

*E-mail: eric.hf.chen@hotmail.com

**E-mail: bogushr@mail.ru

Поступила в редакцию 25.02.2022 г.

После доработки 18.04.2022 г.

Принята к публикации 17.05.2022 г.

Статья посвящена исследованию влияния функции активации в сверточных нейронных сетях на точность реидентификации людей на изображениях, полученных с различных видеокамер распределенных систем видеонаблюдения. Проведены исследования для наиболее применяемых функций активации при обнаружении объектов на изображениях, таких как ReLU, Leaky-ReLU, PReLU, RReLU, ELU, SELU, GELU, Swish, Mish по критериям: точность реидентификации человека с использованием метрик Rank1, Rank5, Rank10, mAP и время, затраченное на обучение модели. В качестве экстрактора признаков использовались архитектуры ResNet-50, DenseNet-121 и DarkNet-53. Экспериментальные исследования выполнены на открытых наборах данных Market1501 и PolReID. Оценка точности реидентификации выполняется после трехкратного повторения обучения и тестирования при использовании разных функций активации, архитектур нейронных сетей и наборов данных с использованием усреднения полученных значений метрик.

DOI: 10.31857/S013234742205003X

1. ВВЕДЕНИЕ

Повторная идентификация (реидентификация) людей является важной задачей компьютерного зрения и позволяет обнаруживать одного и того же человека в разных местах и в разное время на изображениях, полученных с видеокамер распределенных систем видеонаблюдения. Однако реидентификация требует учета многих мешающих факторов включая окклюзии, разный уровень освещенности, различное разрешение видеокамер, отличающийся внешний вид одного и того же человека из-за различных ракурсов его съемки, схожесть внешних признаков между разными людьми. Повторная идентификация схожа

с классификацией, однако существенное отличие заключается в отсутствии определенного количества классов объектов, так как количество разных людей может быть различным.

В настоящее время наиболее эффективными инструментами для решения данной задачи являются сверточные нейронные сети (СНС), которые используются для извлечения признаков на изображении каждого человека. В [1] для алгоритма реидентификации небольшого количества людей в помещениях используют модифицированную быстродействующую СНС на основе ResNet-34. В [2] для реидентификации применяются СНС ResNet-34 и ResNet-50 и показано, что Res-

Net-50 требует больших временных затрат, но обеспечивает увеличение точности.

СНС DenseNet-121 [3] характеризуется наличием соединений между слоями, при которых карты признаков всех предыдущих слоев используются в качестве входных для всех последующих в блоке. Кроме этого, карты признаков не суммируются от слоя к слою, что характерно для ResNet, а конкатенируются. Такие особенности данной СНС позволили повысить эффективность реидентификации по сравнению с ResNet-50 [4, 5].

Архитектура DarkNet-53 была предложена в качестве базовой СНС для извлечения признаков в алгоритме обнаружения объектов YOLOv3. При оценке в метриках top-1 и top-5 на базе данных ImageNet в [6] показано, что данная СНС по точности соизмерима с ResNet-152, а по времени обработки быстрее более чем в два раза. Таким образом, DarkNet-53 является перспективной для повторной идентификации людей.

Известно, что эффективность дескриптора объекта определяется архитектурой СНС и выборкой данных, на которой она была обучена. Увеличение глубины нейронных сетей позволяет повысить точность работы, однако приводит к тому, что на этапе обучения при обратном распространении ошибки могут происходить такие явления, как взрывные или исчезающие градиенты. Взрывные градиенты приводят к проблеме, возникающей при накоплении больших градиентов ошибок, за счет чего веса сети обновляются слишком быстро. Это делает модель нестабильной и неспособной к обучению на тренировочных данных. Исчезающие градиенты представляют обратную проблему, которая так же приводит к невозможности эффективного обучения модели. Обучение СНС методом обратного распространения ошибки предполагает корректировку весов в зависимости от функции потерь и скорости обучения. При этом веса умножаются на производную от функции потерь, и чем глубже сеть, тем меньше эта величина. При значении градиента близком к нулю веса СНС вообще перестанут обновляться. Существуют разные способы решения этих проблем, одним из которых является поиск функции активации (ФА) для решения прикладной задачи. Целью данной статьи является исследование и выбор наиболее эффективной ФА для использования в сверточных нейронных сетях при повторной идентификации людей в системах видеонаблюдения.

2. ФУНКЦИИ АКТИВАЦИИ В СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЯХ

В настоящее время существуют различные ФА, и каждая характеризуется своими недостатками и преимуществами, которые влияют на эф-

фективность их применения для решения прикладной задачи. Однако нет универсальной ФА, подходящей для множества разных задач компьютерного зрения.

С учетом схожести классификации и повторной идентификации объектов в системах видеонаблюдения для ФА СНС при реидентификации можно выделить основные требования:

1. Нелинейность. ФА должна быть нелинейной функцией, т.к. для линейной функции производная является константой и при обучении СНС изменения весов не будет зависеть от входных значений. Кроме этого, комбинация линейных функций так же является линейной, за счет чего утрачиваются преимущества использования многослойных СНС.

2. Схожесть ФА на положительном участке области определения с функцией идентичности. Такой подход позволяет ускорить обучение при небольших значениях градиентов.

3. Нулевой участок на отрицательной части области определения. Такие ФА обладают эффектом “прореживания” нейронов, что позволяет облегчить сеть и ускорить обучение. Ненулевой участок на отрицательной части области определения позволяет снизить количество потерянной информации, при этом сохраняет эффект “прореживания”.

4. Значение производной на положительной части области определения должно быть близко к единице. Производная ФА на всей или на большей части области определения для положительных входных значений нейронов должна принимать значения равные или больше 1, чтобы при обратном распространении ошибки по сети не происходило затухание градиентов из-за постоянного умножения на величину, близкую к 0.

Среди указанных обязательным требованием является нелинейность ФА. Схожесть ФА с функцией идентичности обусловлена ее наклоном в 45° к оси x , что позволяет передавать значения активированных нейронов без значительных изменений. Использование кусочно-заданных функций, где на положительной области определения используется функция идентичности, позволяет снизить вычислительные затраты, и, как следствие, ускорить обучение. Третье требование так же направлено на ускорение обучения за счет добавления эффекта “прореживания”. Решить проблему затухающих градиентов может позволить четвертое требование. Полное соответствие ФА всем требованиям позволит повысить эффективность работы СНС.

Выбор ФА для конкретной прикладной задачи предполагает проведение экспериментальных исследований, которые позволят определить наиболее эффективную по точности и временным затратам.

Одной из наиболее распространенных в настоящее время ФА является ReLU (выпрямленной линейной единицы), графически представленная на рис. 1а [7]. ReLU представляет собой кусочно-заданную функцию:

$$\varphi(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (2.1)$$

где x – входное значение нейрона.

Основное преимущество заключается в низкой вычислительной сложности, которая обусловлена тем, что при прямом проходе по сети положительные значения сохраняются, а отрицательные приравниваются к нулю. При обратном проходе вычислительные затраты так же минимальны, и заключаются только в сравнении значений нейронов с нулем: при отрицательных значениях производная равна нулю, при положительных равна единице. Несмотря на сходство с линейной функцией в области положительных значений, ReLU нелинейная, а значит перспективна для реидентификации людей. Однако значения производной на положительной части области определения могут привести к такой проблеме, как взрывные градиенты. Отрицательная часть области определения функции приводит к существованию эффекта, называемого “прореживание нейронов”. Суть его заключается в том, что нейроны, которые не были активированы изначально никогда не смогут быть активированными, и к ним будут добавляться те нейроны, на входы которых поступали отрицательные значения. Алгоритм градиентного спуска не сможет настраивать веса таких нейронов.

Решением проблемы “прореженных” нейронов может быть ФА Leaky-ReLU [8], или ReLU с утечкой:

$$\varphi(x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \leq 0 \end{cases} \quad (2.2)$$

где α – угловой коэффициент, принимающий небольшие значения, традиционно $\alpha = 0.01$.

ФА Leaky-ReLU (рис. 1б) сохраняет для нейронов небольшие отрицательные значения, позволяя им быть активированными. Это достигается за счет коэффициента α , принимающего небольшие значения, обычно 0.01, на который умножается функция идентичности. Использование коэффициента α при обратном проходе позволяет весам обновляться для отрицательных входных значений, из-за отсутствия нулевого градиента, характерного в подобной ситуации для ReLU. Коэффициент α является гиперпараметром, требующим настройки, что можно отнести к недостаткам использования данной ФА, так как подбор коэффициента требует дополнительных исследований. Кроме этого, проблема взрывных градиентов остается открытой для этой ФА.

В [9] представлены результаты эмпирического исследования, в котором определяется влияние угла наклона отрицательной части функции на задаче классификации изображений при использовании ФА ReLU и Leaky-ReLU, а также их модификаций: параметрической выпрямленной линейной единицы (PReLU) и рандомизированной выпрямленной линейной единицы с утечкой (RReLU). Тестирование выполнено на наборах данных CIFAR-10 и CIFAR-100. В качестве СНС, формирующих признаки людей, используются архитектуры NiN и NDSB Network. В PReLU оптимальное значение коэффициента α подбирается в процессе обучения нейронной сети, а в RReLU этот коэффициент задается случайным образом для каждого слоя. Проведенные исследования показали, что лучшие результаты были получены при использовании PReLU. Однако в этом случае высока вероятность переобучения СНС при использовании небольшого набора данных, поэтому RReLU оказывается более эффективной на практике.

Кроме указанных модификаций, небольшой наклон в отрицательной части области определения функции имеют ФА ELU, SeLU, GeLU, что позволяет предположить эффективность их использования для повторной идентификации людей.

ФА ELU (Exponential Linear Unit) [10] определяется выражением:

$$\varphi(x) = \begin{cases} x, & x \geq 0 \\ \alpha(e^x - 1), & x < 0 \end{cases} \quad (2.3)$$

где $\alpha > 0$ – коэффициент, ограничивающий величину выходных значений на отрицательном участке области определения функции.

ФА ELU имеет ненулевые значения в отрицательной области и описывается логарифмической кривой (рис. 1 в). Такой подход должен позволить достигнуть насыщения нейронов для отрицательных входных значений, и тем самым уменьшить вариативность данных, распространяемых дальше по сети. Форма производной повторяет форму кривой для отрицательных значений нейронов, и имеет нулевые значения только для насыщенных нейронов. Ненулевые значения производной позволяют решить проблему “прореженных” нейронов, характерную для ReLU. К недостаткам можно отнести необходимость подбора коэффициента α и большую вычислительную сложность, чем для предыдущих модификаций функции ReLU.

ФА SELU (Scaled Exponential Linear Unit) (рис. 1г) является масштабированным вариантом ELU и описывается выражением:

$$\varphi(x) = \lambda \begin{cases} x, & x \geq 0 \\ \alpha(e^x - 1), & x < 0 \end{cases} \quad (2.4)$$

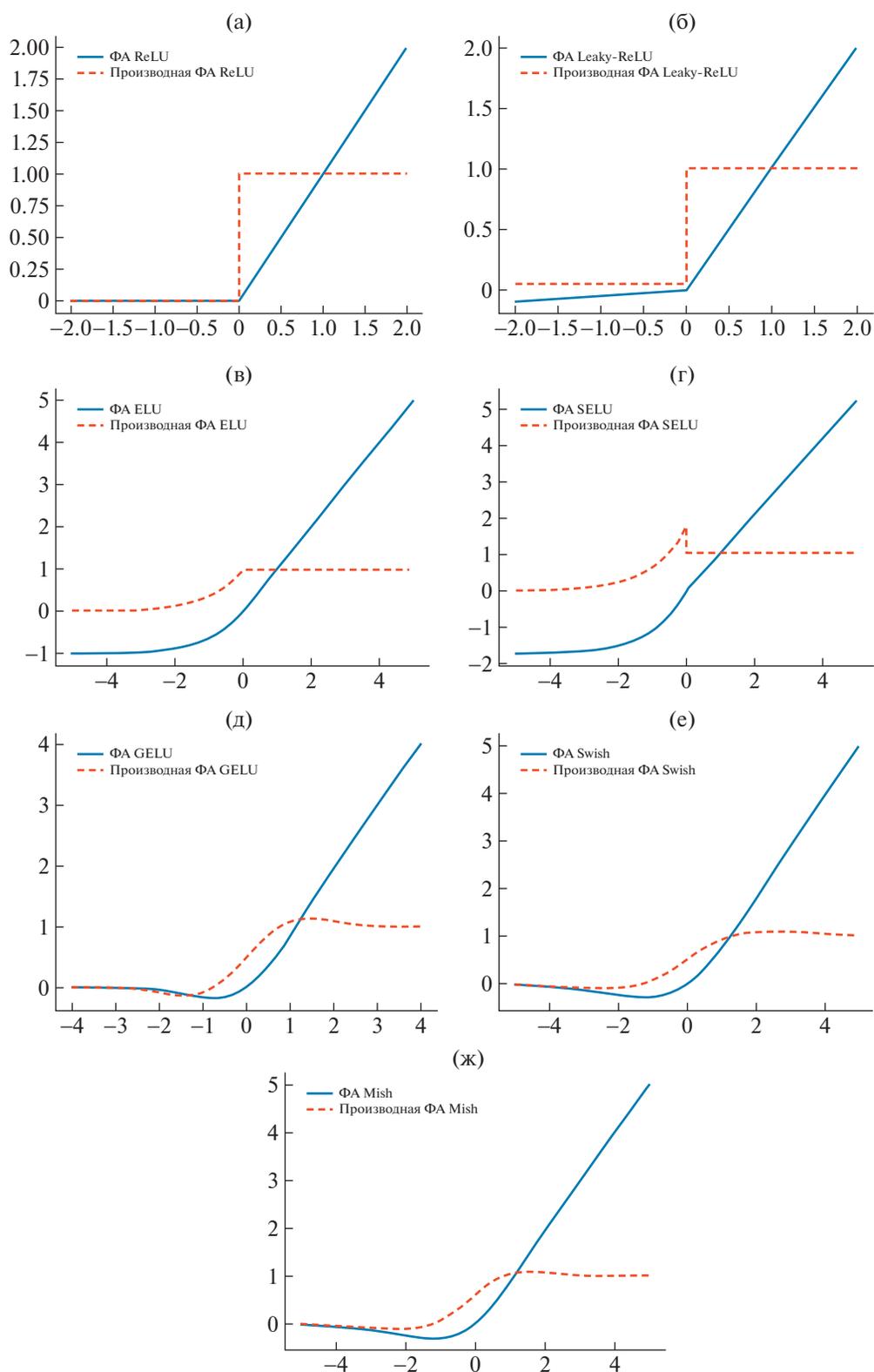


Рис. 1. Графики функций активации и их производной: а – ReLU; б – Leaky-ReLU, где $\alpha = 0.05$; в – ELU, где $\alpha = 1$; г – SELU, где $\alpha = 1.67326$, $\lambda = 1.0507$; д – GELU; е – Swish, где $\beta = 1$; ж – Mish.

В исследовании, представленном в [11], определяются значения для коэффициентов $\alpha = 1.67326$, $\lambda = 1.0507$. Функция SELU обладает эффектом самонормализации, для которого характерно, что среднее значение выходов на каждом слое равно нулю, а стандартное отклонение равно единице. Это позволяет сети иметь более быструю сходимость. Применение SELU предполагает начальную инициализацию весов в соответствии с нормальным распределением для обеспечения самонормализующих свойств.

ФА GELU (Gaussian Error Linear Units) [12] определяется выражением:

$$\begin{aligned} \varphi(x) &= x \cdot \frac{1}{2} [1 + \operatorname{erf}(x/\sqrt{2})] \approx \\ &\approx 0.5x(1 + \tanh(\sqrt{2/\pi}(x + 0.044715x^3))) \end{aligned} \quad (2.5)$$

или:

$$\varphi(x) = x\sigma(1.702x), \quad (2.6)$$

где $\sigma = \frac{1}{1 + e^{-x}}$ – функция активации сигмоиды.

Характерным отличием GELU от рассмотренных выше ФА является то, что GELU это невыпуклая немонотонная функция, нелинейная на всей области определения. Это позволяет GELU легче аппроксимировать сложные функции. В [12] представлено сравнение эффективности ФА ReLU, ELU и GELU для классификации на наборах данных MNIST и CIFAR10/100, и для распознавания речи на наборе данных TIMIT. Для данных задач показано улучшение эффективности их решения при использовании ФА GELU.

В [13] для поиска наиболее эффективной ФА используется подход автоматической генерации, основанный на последовательном переборе унарных и бинарных функций, которые поочередно объединяются, а результат оценивается эмпирически. Тестирование осуществлялось на наборах данных CIFAR-10 и CIFAR-100 с использованием СНС ResNet-164, WRN (Wide ResNet 28-10) и DenseNet 100-12, и на наборе ImageNet с использованием Mobile NASNet-A, Inception-ResNet-v2, Inception-v3 и Inception-v4 для классификации изображений, а также для решения задачи машинного перевода English-German на WMT2014. Наилучший результат при тестировании на разных наборах данных и архитектур нейронных сетей показала функция, получившая название Swish, которая определяется выражением:

$$\varphi(x) = x\sigma(\beta x), \quad (2.7)$$

где β – коэффициент, регулирующий степень кривизны функции, σ – функция сигмоиды.

Swish отличается от GELU коэффициентом β , который можно подбирать для улучшения обобщающей способности сети. Уменьшение β приближает эту функцию к линейной, а увеличение β –

к ReLU. Swish не ограничена сверху, что ускоряет обучение при почти нулевых градиентах. Ограничение снизу для данной ФА приводит к эффекту регуляризации и позволяет отбросить большие отрицательные входные значения, что важно на первых шагах обучения, когда возникают большие отрицательные сигналы.

В [14] рассмотрена ФА, получившая название Mish, которая определяется выражением:

$$\varphi(x) = x \tanh(\operatorname{softplus}(x)) = x \tanh(\ln(1 + e^x)). \quad (2.8)$$

Mish визуально схожа с ФА Swish и GELU (рис. 1д, е, ж), обладает всеми теми же свойствами. В большинстве экспериментов приведенных в [13], Mish показала лучшие результаты, чем другие рассмотренные ФА для классификации объектов на наборах данных CIFAR-10 и MNIST для СНС ResNet-20, WRN-10-2, SimpleNet, Xception Net, Capsule Net, Inception ResNet v2, DenseNet-121, MobileNet-v2, ShuffleNet-v1, Inception v3, EfficientNet B0 на наборе данных ImageNet-1k с СНС ResNet-18, ResNet-50, SpineNet-49, PeleeNet, CSP-ResNet-50, CSP-DarkNet-53, CSP-ResNext-50. Для задачи обнаружения людей на наборе данных MS COCO с СНС CSP-DarkNet-53 и CSP-DarkNet-53+PANet+SPP и в составе архитектуры СНС для обнаружения объектов YOLOv4 в различных модификациях Mish также позволяет улучшить результативность.

3. НАБОРЫ ДАННЫХ ДЛЯ РЕИДЕНТИФИКАЦИИ

Для алгоритмов реидентификации одним из крупных и распространенных наборов данных, с изображениями людей с нескольких камер видеонаблюдения является Market1501 [15]. Этот набор данных содержит 32668 различных изображений для 1501 человека. Для формирования набора данных использовались кадры, полученные с шести камер видеонаблюдения, установленных возле супермаркета в университете Циньхуа. Набор данных разделен на обучающую и тестовую выборки, и для обучения используются 12936 изображений для 751 человека, а в галерею включены и используются для тестирования 19732 изображения для 750 человек. Набор данных так же содержит 2793 ограничивающих прямоугольника, содержащих неверные, но достаточно правдоподобные ответы, включенные в тестовую выборку.

Для повышения объективности оценки точности реидентификации с применением разных ФА предложен и используется другой набор данных PolReID [16], сформированный при участии сотрудников и студентов Полоцкого государственного университета, которые подтвердили свое согласие на применение их изображений для проведения экспериментов. Таким образом, PolReID

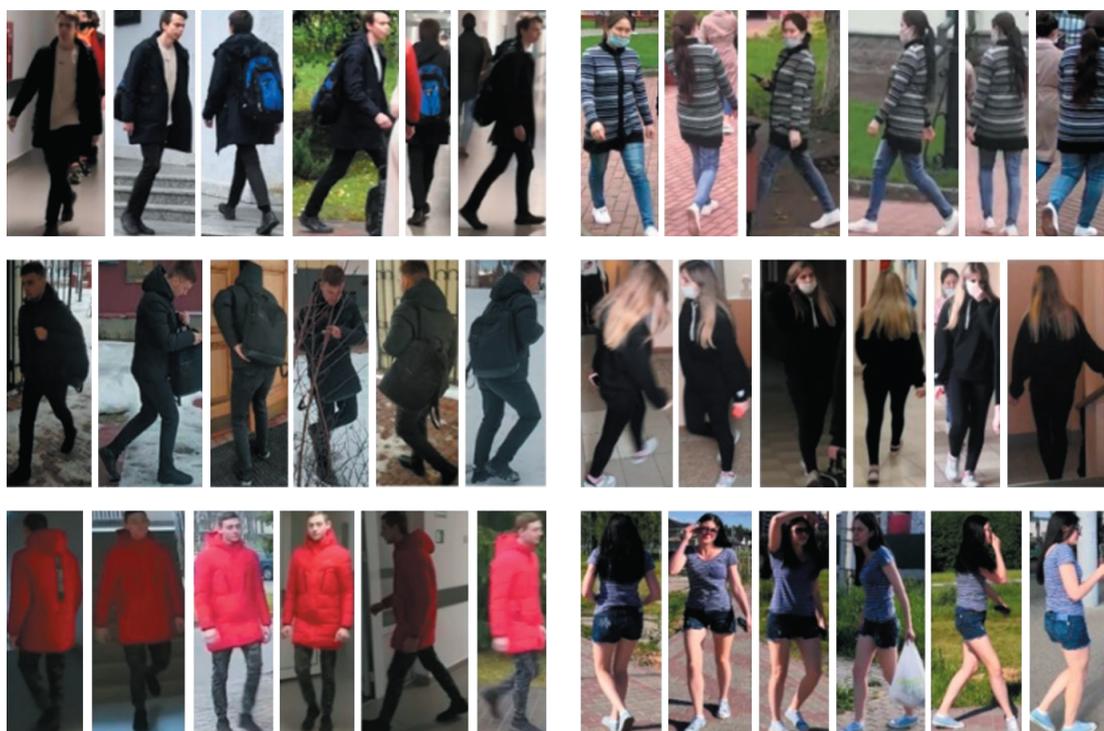


Рис. 2. Примеры изображений набора данных PolReID.

является открытым набором данных и может использоваться другими исследователями [17]. Необходимость тестирования на нескольких наборах данных обусловлена проблемой смещения доменов при реидентификации. При обучении СНС на одном наборе данных, а тестировании на данных, полученных в других условиях, точность реидентификации значительно снижается, но является более объективной на практике, чем при использовании выборки, принадлежащей одному домену. В [16] показано, что объединение нескольких наборов данных позволяет повысить точность реидентификации в целом.

Текущий набор PolReID включает 31919 изображений для 271 человека, при этом обучающая выборка представлена 16770 изображениями для 145 человек, а тестовая – 15149 изображений для 126 человек. Изображения людей из кадров получены с помощью СНС YOLOv4 [18] и рассортированы оператором. Для формирования комплекта изображений для каждого человека использовалось от двух до десяти камер и от одной до девяти видеопоследовательности с каждой камеры. Таким образом, всего использовалось более чем 200 локаций съемки, среди них в уличных условиях 159 человек, 40 из которых также зафиксированы камерами внутреннего наблюдения в помещениях с естественным и искусственным освещением различной интенсивности. Изображения получены при разных погодных условиях и временах го-

да: летом для 47 человек, осенью для 176 человек, зимой для 48 человек. PolReID включает изображения 194 мужчин и 77 женщин разной возрастной категории: 230 человек в возрасте до 30 лет, 41 человек в возрасте от 30 до 60 лет. Причем 134 человека на лице имеют маску, 9 из них на некоторых кадрах присутствуют без маски. Изображения людей представлены с нескольких ракурсов с различными предметами, такими как сумка, пакет, телефон, рюкзак, очки. Несколько примеров изображений из PolReID показаны на рис. 2.

При объединении наборов данных Market1501 и PolReID для корректной работы алгоритма реидентификации все имена изображений были приведены к единому формату записи: XXXXX_sYYYsZZ_AAAAAA_BB.jpg, где XXXXX – идентификатор человека, YYY – номер камеры, ZZ – номер видеопоследовательности, полученный с этой камеры, AAAAAA – номер кадра в видеопоследовательности, BB – количество человек на текущем кадре.

Таким образом, объединенный набор данных всего содержит 68587 ограничительных рамок для 1772 человек. Для обучения использовалось 29707 изображений для 896 человек, а для тестирования 38881 изображение для 876 человек.

Таблица 1. Результаты экспериментов по определению наиболее эффективных функций активации для реидентификации человека на наборе данных Market1501

Функция активации	ResNet-50			DenseNet-121			DarkNet-53		
	Время обучения (м:с)	Rank1	mAP	Время обучения (м:с)	Rank1	mAP	Время обучения (м:с)	Rank1	mAP
ReLU	86:55	81.24*	57.31*	92:06	79.63*	56.62*	98:08	80.08	57.12
LeakyReLU	86:50	81.50	57.44	95:52	79.84	56.68	99:31	80.97*	57.16*
PReLU	102:49	79.57	55.54	138:25	78.12	53.82	112:09	76.93	52.87
RReLU	90:51	81.00	58.24	101:57	79.36	56.68	110:00	79.96	56.73
ELU	85:26	79.51	56.17	99:54	73.52	49.09	111:16	72.98	46.72
SELU	86:21	77.46	51.91	95:12	66.57	40.73	97:35	66.09	40.31
GELU	88:08	81.59	57.92	95:13	79.69	56.98	107:53	80.98	57.26
Swish	104:42	81.38	57.68	128:57	80.08	56.87	116:10	81.09	58.65
Mish	104:52	82.16	58.95	130:13	80.76	57.10	117:48	81.44	57.97

* – отмечены значения Rank1 и mAP, полученные для базовой функции активации для каждой из рассматриваемых СНС

4. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

4.1. Метрики для оценки точности реидентификации

Для оценки точности работы алгоритма реидентификации использовались метрики Rank1, Rank5, Rank10 и mAP.

RankN является метрикой качества ранжирования и показывает процент числа запросов, для которых верный выданный результат был среди первых N полученных результатов. Соответственно, метрика Rank1 показывает процент запросов, для которых идентификатор первого изображения-кандидата совпадает с идентификатором запроса. Rank5 показывает процент запросов, для которых среди первых пяти выданных изображений-кандидатов было верное решение, Rank 10 – среди первых 10. Для вычисления RankN определяется отношение суммы запросов, для которых верное решение было найдено среди первых выданных результатов, к общему числу запросов Q:

$$RankN = \frac{\sum K_{i,N}}{Q}, \quad (2.9)$$

где: i – номер запроса; $K_{i,N}$ – i-тый запрос, для которого верное решение было найдено среди первых N выданных результатов.

Метрика mAP является оценкой точности алгоритма повторной идентификации, отражающей среднее значение средней точности для всех запросов и рассчитывается по формуле:

$$mAP = \frac{1}{Q} \sum_{i=1}^Q AP_i, \quad (2.10)$$

где Q – общее число запросов, $AP = \frac{\sum precision}{I}$ – средняя точность для каждого i-того запроса, I – число изображений в тестовой выборке, $precision = \frac{TP}{TP + FN}$ – точность запроса, TP – количество верно положительных предсказаний запроса, FN – ложно-отрицательных предсказаний запроса.

4.2. Результаты экспериментов

Первый этап экспериментов направлен на выявление наиболее эффективных ФА для повторной идентификации с использованием алгоритма [19], реализованном на фреймворке pyTorch. Обучение модели выполнено для ResNet-50, DenseNet-121 и DarkNet-53 без применения предварительного обучения, в течение 60 эпох со скоростью 0.03, уменьшенной после 40-й эпохи на 0.01, с размером пакета 16, на наборе данных Market1501 с использованием персонального компьютера с основными характеристиками: Intel Core i5 3.11 GHz, 16 Gb RAM, Nvidia GeForce RTX-3060 6 Gb.

В табл. 1 представлены результаты тестирования алгоритма реидентификации человека с различными ФА в СНС и отмечены значения метрик точности реидентификации, которые обеспечивают повышение значений по сравнению с базовой ФА: СНС ResNet-50 и DenseNet-121 реализации используют ФА ReLU, в DarkNet-53 используется Leaky-ReLU.

Из табл. 1 очевидно, что ФА PReLU, RReLU, ELU и SELU для задачи реидентификации наименее эффективны, точность реидентификации

Таблица 2. Влияние функции активации в СНС ResNet-50 на точность реидентификации людей при обучении на объединенном наборе данных Market1501 и PolReID

Функция активации	Market1501		PolReID		Market1501 и PolReID	
	Rank1	mAP	Rank1	mAP	Rank1	mAP
ReLU*	83.82	62.35	86.60	64.11	84.23	62.20
	84.00	62.24	88.55	63.23	84.62	62.00
	83.49	62.15	87.95	63.09	84.10	61.92
	83.77	62.25	87.70	63.48	84.32	62.04
Leaky-ReLU	83.25	62.50	86.90	63.85	83.76	62.30
	84.00	62.56	87.20	63.06	84.50	62.28
	84.12	63.21	88.40	63.10	84.62	62.78
	83.79	62.76	87.50	63.34	84.29	62.45
GELU	84.53	62.71	86.90	63.99	84.85	62.61
	83.76	62.50	89.76	64.28	84.72	62.46
	83.82	62.43	87.05	63.39	84.15	62.21
	84.04	62.55	87.90	63.89	84.57	62.43
Swish	83.37	61.08	88.86	63.96	84.20	61.25
	83.31	61.52	91.11	65.33	84.47	61.81
	83.91	61.51	89.46	65.33	84.72	61.79
	83.53	61.37	89.81	64.87	84.46	61.62
Mish	83.94	62.37	88.55	64.75	84.62	62.42
	82.81	62.10	88.55	65.09	83.58	62.24
	83.52	62.32	88.10	65.70	84.23	62.53
	83.42	62.26	88.40	65.18	84.14	62.40

* – отмечена базовая функция активации для СНС ResNet-50

при их использовании значительно меньше, чем при использовании базовой функции активации СНС. При применении ФА PReLU наблюдается снижение метрики Rank1 для всех рассмотренных. Максимальное уменьшение mAP характерно для СНС DarkNet-53 и составляет 4.29%. При использовании RReLU следует отметить снижение точности Rank1 по сравнению с базовой ФА. Применение ФА ELU и SELU приводит к значительному уменьшению значений Rank1 и mAP для СНС DenseNet-121 и DarkNet-53. Для всех используемых СНС наилучшие значения Rank1 и mAP характерны для ФА GELU, Swish и Mish, значит требуются дальнейшие исследования с расширением используемой выборки данных.

С учетом полученных результатов, на втором этапе исключены наименее эффективные ФА.

Для оценки воспроизводимости результатов точности повторной идентификации людей каждая СНС с разными ФА была повторно обучена

трижды, каждый раз с одинаковыми гиперпараметрами на объединенном наборе данных Market1501 и PolReID.

Тестирование выполнялось на тестовых выборках отдельно для Market1501, PolReID и на объединенном наборе для каждой обученной СНС с последующим определением среднего арифметического значений метрик. Набор данных Market1501 разбит на тестовую и обучающую выборки согласно протоколам исходных документов, PolReID разделен случайным образом на обучающую (53%) и тестовую выборку (47%), которые не пересекаются. В табл. 2–4 представлены результаты тестирования реидентификации с использованием СНС ResNet-50, DenseNet-121 и DarkNet-53.

Строки в табл. 2–4, выделенные серым цветом содержат средние значения для каждой СНС. Числовые значения метрик, отмеченные полужирным шрифтом, являются лучшими среди

Таблица 3. Влияние функции активации в СНС DenseNet-121 на точность реидентификации людей при обучении на объединенном наборе данных Market1501 и PolReID

Функция активации	Market1501		PolReID		Market1501 и PolReID	
	Rank1	mAP	Rank1	mAP	Rank1	mAP
ReLU*	83.19	62.27	88.40	63.46	83.83	62.11
	83.76	62.22	88.70	63.35	84.57	62.09
	82.57	62.08	87.50	63.37	83.31	61.95
	83.17	62.19	88.20	63.39	83.90	62.05
Leaky-ReLU	82.99	62.14	87.65	63.22	83.73	62.00
	83.94	61.91	88.10	64.79	84.55	62.03
	83.34	62.17	87.80	63.72	83.98	62.06
	83.42	62.07	87.85	63.91	84.09	62.03
GELU	82.81	61.09	89.76	64.60	83.85	61.30
	82.63	60.95	89.16	64.93	83.48	61.23
	82.63	60.62	90.21	63.70	83.76	60.76
	82.69	60.89	89.71	64.41	83.70	61.10
Swish	81.65	58.72	88.70	63.48	82.76	59.14
	81.56	58.35	88.86	63.12	82.64	58.72
	80.76	58.22	88.40	62.08	81.97	58.52
	81.32	58.43	88.65	62.89	82.46	58.79
Mish	82.04	59.14	88.40	62.35	83.08	59.33
	80.76	59.07	87.95	62.04	81.87	59.20
	82.54	59.46	87.85	63.22	83.36	59.71
	81.78	59.22	88.07	62.54	82.77	59.41

* – отмечена базовая функция активации для СНС DenseNet-121

средних для каждой тестовой выборки. Анализ табл. 2–4 показывает, что при повторении экспериментов для СНС, не изменяя ее архитектуру гиперпараметры, обучающую и тестовую выборки возможен существенный разброс значений метрик точности реидентификации. Так, для СНС

DarkNet-53 с ФА Swish на тестовой выборке Market1501 разница между лучшим и худшим результатом mAP составляет 3.26 (см. табл. 4). Поэтому оценка средних значений метрик является предпочтительнее. Анализ средних значений метрик моделей СНС показывает, что при разных тестовых

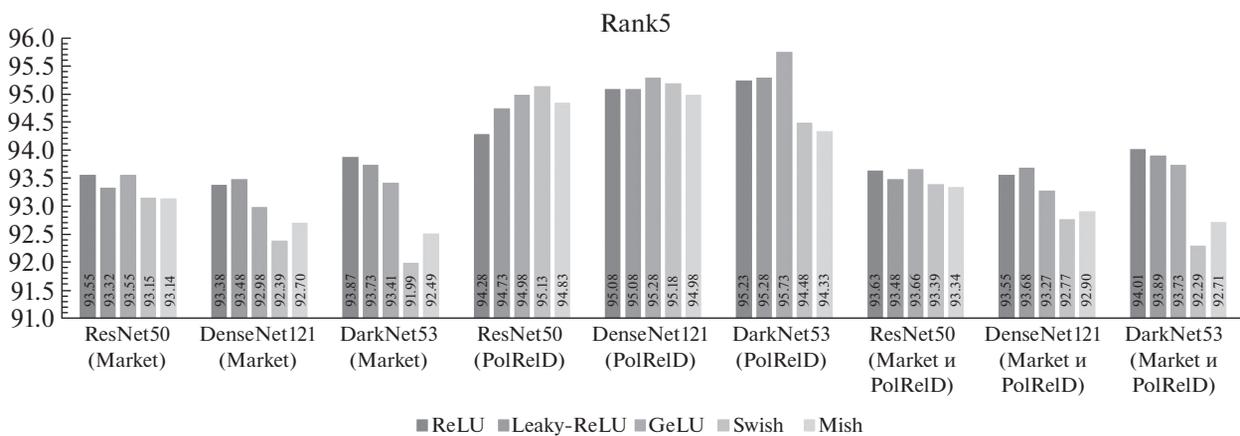


Рис. 3. Значения метрики Rank5 для различных СНС, ФА и наборов данных.

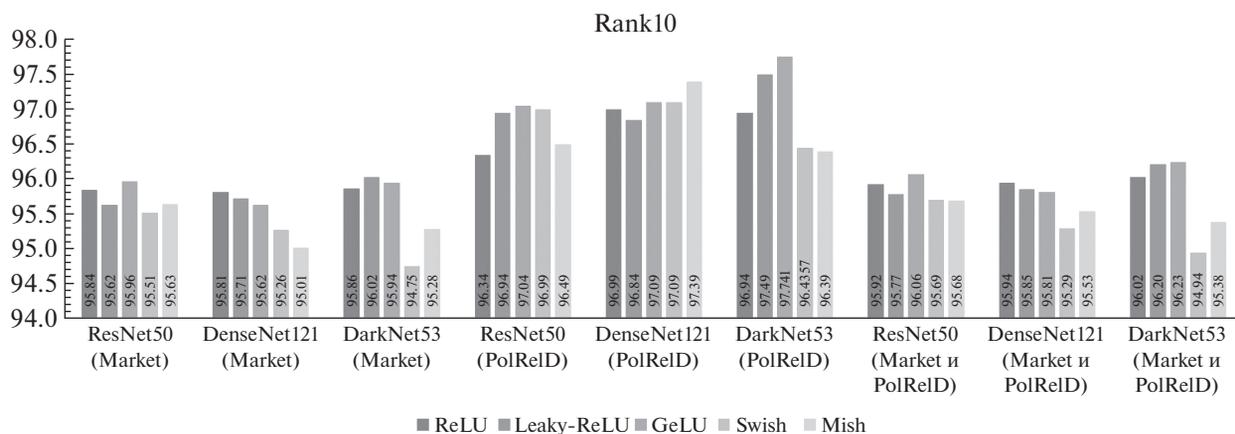
Таблица 4. Влияние функции активации в СНС DarkNet-53 на точность реидентификации людей при обучении на объединенном наборе данных Market1501 и PolReID

Функция активации	Market1501		PolReID		Market1501 и PolReID	
	Rank1	mAP	Rank1	mAP	Rank1	mAP
ReLU	83.91	63.61	88.40	64.80	84.55	63.47
	84.32	63.02	87.20	64.28	84.70	62.84
	84.59	63.63	89.16	65.27	85.27	63.50
	84.27	63.42	88.25	64.78	84.84	63.27
Leaky-ReLU*	83.08	62.43	87.95	64.45	83.68	62.40
	83.61	63.00	89.16	64.94	84.45	62.97
	84.95	64.03	89.46	65.39	85.59	63.89
	83.88	63.16	88.86	64.93	84.57	63.09
GELU	82.22	60.11	87.50	62.93	82.86	60.13
	84.00	62.28	89.91	64.84	84.90	62.34
	83.79	62.25	89.46	65.08	84.67	62.37
	83.34	61.55	88.96	64.28	84.14	61.61
Swish	80.82	55.56	87.95	59.80	81.87	55.75
	81.41	58.82	88.70	62.55	82.34	58.98
	80.43	55.90	87.35	60.63	81.47	56.22
	80.89	56.76	88.00	60.99	82.61	56.98
Mish	81.80	57.85	87.50	61.56	82.61	58.02
	81.89	58.98	88.86	62.35	82.99	59.11
	80.43	56.96	86.30	60.74	81.13	57.11
	81.37	57.93	87.55	61.55	82.24	58.08

* – отмечена базовая функция активации для СНС DarkNet-53

вых выборках выделить одну ФА, наиболее эффективную для всех СНС не представляется возможным, так как лучшие показатели имеют разные ФА. Из табл. 2–4 очевидно, что в большинстве

случаев наиболее эффективна функция ReLU, для которой средние значения Rank1 более значимы в трех экспериментах, mAP наилучшие в четырех экспериментах. Функция GeLU обеспе-

**Рис. 4.** Значения метрики Rank10 для различных СНС, ФА и наборов данных.

чивает максимизацию показателей для Rank1 в четырех случаях, а для mAP в одном.

На рис. 3 и рис. 4 показано сравнение значений метрик Rank5 и Rank10, соответственно, при обучении на объединенном наборе данных и тестировании на Market1501, PolReID и объединенном наборе данных Market1501 и PolReID для различных ФА.

Рис. 3 свидетельствует о том, что при оценке алгоритма реидентификации в метрике Rank5 в трех случаях из девяти наилучшие показатели характерны для ФА GeLU, в двух случаях для ReLU. Следует отметить, что применение ФА ReLU и GeLU в СНС ResNet-50 при тестировании на наборе данных Market1501 приводит к одинаковым средним значениям Rank5.

Анализ рис. 4, отражающего точность реидентификации в метрике Rank10 показал, что ФА GeLU обеспечивает более высокие показатели для большинства рассмотренных случаев.

5. ОБСУЖДЕНИЕ И РЕКОМЕНДАЦИИ

Анализ проведенных исследований показывает, что при оценке точности реидентификации с использованием метрик Rank5 и Rank10 наиболее эффективны ФА GeLU и ReLU. Причем GeLU показала лучший результат в трех случаях из девяти, а ReLU имела наибольшие результаты в двух экспериментах для Rank5. Сравнение точности по метрике Rank10 свидетельствует о том, что ФА GeLU показала наилучшие результаты в пяти случаях из девяти.

Оценка по метрикам Rank1 и mAP позволяет выделить наиболее эффективную ФА ReLU.

Разброс значений метрик точности реидентификации при повторном обучении СНС без изменения ее архитектуры гиперпараметров, ФА, обучающей и тестовой выборки варьируется для разных СНС и достигает максимального для DarkNet-53 с ФА Swish. Наименьший разброс значений зафиксирован для СНС ResNet-50 с ФА Leaky-ReLU. Однозначно сложно утверждать, что какая-то из рассмотренных ФА обладает большей стабильностью, так как пока неизвестна взаимосвязь между архитектурой СНС, ФА, набором данных и величиной разброса полученных значений точности для Rank1 и mAP. Однако анализ результатов свидетельствует о том, что для разных тестовых выборок и СНС лучшие средние показатели в большинстве случаев имеет ФА ReLU, что говорит о большей воспроизводимости результатов, по сравнению с другими ФА.

Усреднение показателей точности реидентификации для разных наборов данных и архитектур СНС позволило установить, что наиболее целесообразным является использование ФА ReLU для повторной идентификации, она показывает

наилучшее соотношение точность реидентификации—скорость обучения—воспроизводимость результатов для разных моделей СНС.

6. ЗАКЛЮЧЕНИЕ

В работе рассмотрены и проанализированы наиболее распространенные функции активации, используемые в сверточных нейронных сетях, в приложении к задаче повторной идентификации человека в распределенных системах видеонаблюдения. При этом использованы три разные архитектуры СНС, ФА ReLU, Leaky-ReLU, PReLU, RReLU, ELU, SELU, GELU, Swish, Mish. Предложен набор данных для тестирования алгоритмов повторной идентификации PolReID, который в настоящее время включает 31 919 изображений для 271 человека. В результате анализа полученных результатов, было установлено, что для задачи реидентификации наиболее перспективными ФА являются ReLU и GeLU, при этом скорость работы и воспроизводимость результатов для ФА ReLU выше, чем при использовании GeLU.

БЛАГОДАРНОСТЬ

Эта работа поддержана Национальной программой высококвалифицированных иностранных экспертов (Гранты G2021016028L и G2021016001L) и Специальным фондом фундаментальных научных исследований Чжэцзянского Шурен Университета.

СПИСОК ЛИТЕРАТУРЫ

1. *Ye S., Bohush R.P., Chen H. et al.* Person Tracking and Reidentification for Multicamera Indoor Video Surveillance Systems // *Pattern Recognit. Image Anal.* 2020. V. 30. P. 827—837. <https://doi.org/10.1134/S1054661820040136>
2. *Porrello A., Bergamini L., Calderara S.* Robust Re-Identification by Multiple Views Knowledge Distillation. // *ArXiv, abs/2007.04174.* 2020.
3. *Huang G., Liu Z., Weinberger K.Q.* Densely Connected Convolutional Networks. // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2017. P. 2261—2269.
4. *Wang G., Lai J., Huang P., Xie X.* // *Spatial-Temporal Person Re-identification.* // *ArXiv, abs/1812.03282.* 2019.
5. *Mao S., Zhang S., Yang M.* // *Resolution-invariant Person Re-Identification.* // *ArXiv, abs/1906.09748.* 2019.
6. *Redmon J., Farhadi A.* // *YOLOv3: An Incremental Improvement.* *ArXiv abs/1804.02767.* 2018
7. *Nair, Vinod, Geoffrey E. Hinton* Rectified linear units improve restricted Boltzmann machines. // *In ICML.* 2010. P. 807—814.

8. *Maas, L. Andrew*. Rectifier non linearities improve neural network acoustic models. In ICML. 2013. V. 30.
9. *Xu B., Wang N., Chen T., Li M.* Empirical Evaluation of Rectified Activations in Convolutional Network. // ArXiv, abs/1505.00853. 2015.
10. *Clevert D., Unterthiner T., Hochreiter S.* Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) // arXiv: abs/1511.07289v5. 2016.
11. *Klambauer G., Unterthiner T., Mayr A., Hochreiter S.* Self-Normalizing Neural Networks. // ArXiv, abs/1706.02515. 2017.
12. *Hendrycks D., Gimpel K.* Bridging Nonlinearities and Stochastic Regularizers with Gaussian Error Linear Units. // ArXiv, abs/1606.08415. 2016.
13. *Ramachandran P., Zoph B., Le Q.V.* Swish: a Self-Gated Activation Function. // arXiv: abs/1710.05941v2. 2017.
14. *Misra D.* Mish: A Self Regularized Non-Monotonic Neural Activation Function. // ArXiv, abs/1908.08681. 2019.
15. *Zheng L., Shen L., Tian L., Wang Sh., Wang J., Tian Q.* Scalable Person Re-Identification: A Benchmark // IEEE International Conference on Computer Vision (ICCV2015). 2015. P. 1116–1124.
16. *Ihnatsyeva S., Bohush R., Ablameyko S.* Joint Dataset for CNN-based Person Re-identification // Pattern Recognition and Information Processing (PRIP'2021): Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. Minsk: UIIP NASB. 2021. P. 33–37.
17. *Ihnatsyeva S., Bohush R.* PolReID. <https://github.com/SvetlanaIgn/PolReID>. 2021.
18. *Bochkovskiy A., Wang Ch.-Y., Liao H.-Y.M.* YOLOv4: Optimal Speed and Accuracy of Object Detection // ArXiv, abs/2004.10934. 2020.
19. Person reID baseline pytorch. https://github.com/layumi/Person_reID_baseline_pytorch.