

Исследование конфигураций нейросетевых кодеков для адаптивной системы сжатия кадров FPV-видеопотока при управлении беспилотными системами. Часть I. Методика

А.А. Березкин, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича (СПбГУТ), доцент, к.т.н.; berezkin.aa@sut.ru

А.А. Ченский, СПбГУТ, магистрант; chenskii.aa@sut.ru

Р.В. Киричек, СПбГУТ, ректор, профессор, д.т.н.; kirichek@sut.ru

А.А. Захаров, ФГБУ НИИР, заместитель генерального директора по науке, доцент, к.т.н.; a.zaharov@niir.ru

УДК 004.622+004.021

DOI: 10.34832/ELSV.2024.58.9.004

Аннотация. В данной работе исследованы различные конфигурации нейросетевых кодеков, разработанных на основе вариационных автокодировщиков вида VQ и KL (VQ-f4, VQ-f8, VQ-f16, KL-f4, KL-f8, KL-f16, KL-f32) из состава моделей Stable Diffusion на предмет возможности их использования для сжатия кадров видеопотока при FPV-управлении беспилотными системами. В первой части статьи обобщаются результаты исследований сжатия и квантования латентных пространств вариационных автокодировщиков, представляется модификация нейросетевого кодека, наборы данных и рассматривается методика экспериментов.

Ключевые слова: вариационный автокодировщик, квантование, сжатие данных, нейронные сети, латентное пространство признаков, сжатие видеопотока, FPV-управление, БПЛА.

ВВЕДЕНИЕ

На текущий момент беспилотные системы (БС) широко распространены в гражданской и военной сферах [1–3]. В ряде случаев необходимо использовать FPV-управление, представляющее собой удалённое управление БС оператором [1–3]. Для обеспечения канала связи между БС и оператором на больших расстояниях могут использоваться мобильные сотовые сети связи, однако область покрытия таких сетей достаточно ограничена. Для контроля и управления БС в труднодоступных районах в перспективе могут быть использованы гибридные орбитально-наземные сети связи (ГОНСС) [4]. Для эффективного использования каналов передачи видеопотока через ГОНСС необходимо обеспечить высокую степень сжатия передаваемых кадров. Такое сжатие требует использования альтернативных кодеков, в том числе на базе искусственного интеллекта. Одним из таких подходов является использование нейросетевых автокодировщиков, преобразующих данные кадра видеопотока из пиксельного пространства во внутреннее латентное пространство признаков (ЛПП) меньшей размерности.

Метод использования нейросетевых автокодировщиков (прежде всего, вариационных) широко

представлен в научной литературе [1–3, 5–16]. При этом работа кодека видеопотока состоит из следующих этапов: кодирование поступившего кадра в ЛПП; квантование ЛПП для уменьшения размера; дополнительное сжатие квантованного ЛПП методами сжатия с потерями или без потерь; передача кадра от БС на станцию оператора; декомпрессия и деквантование ЛПП; декодирование кадра в пиксельное пространство. В [15, 16] для сжатия FPV-видеопотока используется диффузионная нейросетевая модель и исключение шумов с помощью сверточной нейронной сети U-Net [17]. Также в серии исследований используется интерполяция кадров до разрешения HD [1–3] либо Full HD [15, 16].

В исследованиях [1–3] предложено использование вариационного автокодировщика VQ-f16 вместо диффузионных моделей [15, 16] и определен лучший способ обработки ЛПП, где под обработкой ЛПП понимается сочетание алгоритмов сжатия и квантования ЛПП.

В [1] рассмотрены алгоритмы сжатия без потерь и предложен метод использования форматов изображений для сжатия с потерями путём приведения ЛПП вариационного автокодировщика VQ-f16 размерностью (1, 8, 32, 32) к одноканальному изображе-

нию размерности (1, 128, 64). В [1] представлена разработанная методика сравнения различных методов обработки ЛПП, использующая часть показателей эффективности функционирования каналов информационного обмена (КИО) беспилотных систем [18]. В результате исследования было показано, что с линейным квантованием лучшим алгоритмом сжатия без потерь является алгоритм LZMA (Lempel-Ziv-Markov chain-Algorithm), а сжатия с потерями – AVIF (AV1 Image File Format) с параметром *quality*=45. Кроме того, в [1] была показана неэффективность использования алгоритмов LZW, LZF, LZ4H5, LZ4 и LZ4F и низкая эффективность (степень сжатия (compression ratio) $CR < 5\%$) использования алгоритмов JPEG LS, AEC, BZip2 и LZFSE для сжатия ЛПП.

В [2, 3] рассмотрены алгоритмы квантования и соответствующие им алгоритмы деквантования. Помимо исследованных в [19, 20] линейного (ЛИН), степенного (СТ) и модифицированного логистического алгоритмов (мЛОГ), а также указанного логистического алгоритма (ЛОГ) были предложены новые алгоритмы квантования: гиперболическо-тангенциальный (ГТ), модифицированный гиперболическо-тангенциальный (мГТ), нечётно-степенные (со степенью 3 – СТ3, со степенью 5 – СТ5), двойной логистический (2ЛОГ), модифицированный двойной логистический (м2ЛОГ), гиперболическо-синусоидальный (ГС). Данные алгоритмы – алгоритмы основного квантования [2], целью которых является перевод ЛПП из типа данных float16 (или float32) в uint8. Соответствующие им алгоритмы деквантования осуществляют обратное преобразование из uint8 в float16/float32. Также были определены алгоритмы предварительного квантования для предварительного уменьшения числа уникальных значений с целью увеличения степени сжатия ЛПП [2]. Для этого были исследованы алгоритмы Scale, BitRound, Granular BitRound и BitGroom, имеющие входной параметр *nsd* – число знаков после запятой. Была также модифицирована методика сравнения методов обработки ЛПП, предложенная в [1]. В [2, 3] был получен ряд важных выводов:

1. наилучшим алгоритмом квантования (т.е. сочетание алгоритмов основного и предварительного квантования, а также *nsd*) при использовании алгоритма сжатия без потерь LZMA является мГТ Scale 0;
2. использование *nsd* > 1 (больше одного знака после запятой) нецелесообразно ввиду малого влияния на степень сжатия ЛПП и метрики качества восстанавливаемого кадра;
3. лучшие результаты дают алгоритмы Scale (по степени сжатия) и BitRound (по метрикам качества восстанавливаемого кадра);
4. алгоритмы 2ЛОГ, м2ЛОГ, СТ3 и СТ5 непригодны для использования по причине значительного ухудшения качества восстанавливаемого кадра.

Обобщением исследований [1–3] методов обработки ЛПП вариационного автокодировщика VQ-f16 является совместное исследование алгоритмов сжатия и квантования, в ходе которого было проведено более 4500 экспериментов. Разработана методика ускоренного проведения экспериментов с уточнением части полученных результатов, которая используется далее и в настоящей работе. Было показано, что методы обработки ЛПП с алгоритмами сжатия без потерь менее эффективны тех, где используются алгоритмы сжатия с потерями. Согласно исследованиям [2, 3], в случае алгоритмов сжатия с потерями предварительное квантование приводит к уменьшению качества восстанавливаемого кадра при отсутствии значительного влияния на степень сжатия, а алгоритмы основного квантования мГТ и ГТ при этом не могут эффективно использоваться с алгоритмами сжатия с потерями из-за значительных потерь в качестве восстанавливаемых кадров видеопотока. В результате было определено, что существует целая группа методов сжатия ЛПП, которые показывают близкие результаты по степени сжатия и качеству восстанавливаемого кадра, однако среди них нет явного фаворита при использовании на различных кадрах видеопотока. Это обусловлено влиянием конкретных кадров на степень сжатия кадров некоторыми алгоритмами [21, 22].

Настоящая работа является продолжением исследования алгоритмов сжатия и квантования. Ее целью является обобщение полученных для VQ-f16 результатов на другие предобученные вариационные автокодировщики из моделей [23] Stable Diffusion [24] (VQ-f4, VQ-f8, KL-f4, KL-f8, KL-f16, KL-f32), а также разработка предварительных конфигураций нейросетевого кодера в рамках адаптивной системы сжатия видеопотока [25].

Исследование [25] уделяет внимание контролю интенсивности видеопотока путём предсказания части кадров. В нём рассматриваются две основные задачи: своевременное выявление ухудшения сигнала и включение алгоритмов коррекции; контроль полноты восстановления и решение об информировании оператора о некорректируемом ухудшении качества сигнала и способы его компенсации.

Тем не менее в данных исследованиях не рассматривается проблема увеличения сетевых задержек и потерь пакетов по мере увеличения полезной нагрузки при передаче данных в КИО. Согласно предложенной в [25] схеме, должны существовать различные режимы сжатия кадров видеопотока, которые должны использоваться в зависимости от пропускной способности канала связи между БС и оператором. Реализация режимов и их переключение должны осуществляться на стороне оператора при помощи связки КИО и контроллера режимов, а на стороне БС – при помощи контроллера режимов.

Основой данных режимов является использование лучших методов сжатия кадров видеопотока для определенной пропускной способности КИО, удовлетворяющих требованиям по степени сжатия и качества кадров для каждого режима.

Вследствие отсутствия на текущий момент требований к конкретным режимам функционирования КИО определение лучших методов сжатия кадров видеопотока затруднено. Тем не менее, в данном исследовании выполняется попытка найти некоторые возможные конфигурации нейросетевого кодера. ЛПП различных вариационных автокодировщиков видов VQ и KL обладает различным числом значений и, соответственно, содержит разное количество информации. Максимальные возможные метрики качества восстанавливаемого кадра при применении вариационных автокодировщиков с большим количеством информации в ЛПП выше, а минимальная возможная степень сжатия кадров видеопотока при применении вариационных автокодировщиков с меньшим количеством информации — ниже.

Гипотеза настоящего исследования заключается в том, что лучшие методы сжатия кадров видеопотока для некоторых режимов могут, в зависимости от требований, достигаться при применении вариационных автокодировщиков, имеющих различный размер ЛПП.

Если данная гипотеза верна, то оценка возможных методов сжатия кадров видеопотока для режимов может быть получена путём нахождения лучших методов обработки ЛПП для вариационных автокодировщиков с различными ограничениями по среднему размеру сжатого представления кадров. При различных конфигурациях бинарная последовательность кадров, кодируемых нейросетевым кодером, будет обладать

Рисунок 1
Модель нейросетевого кодера

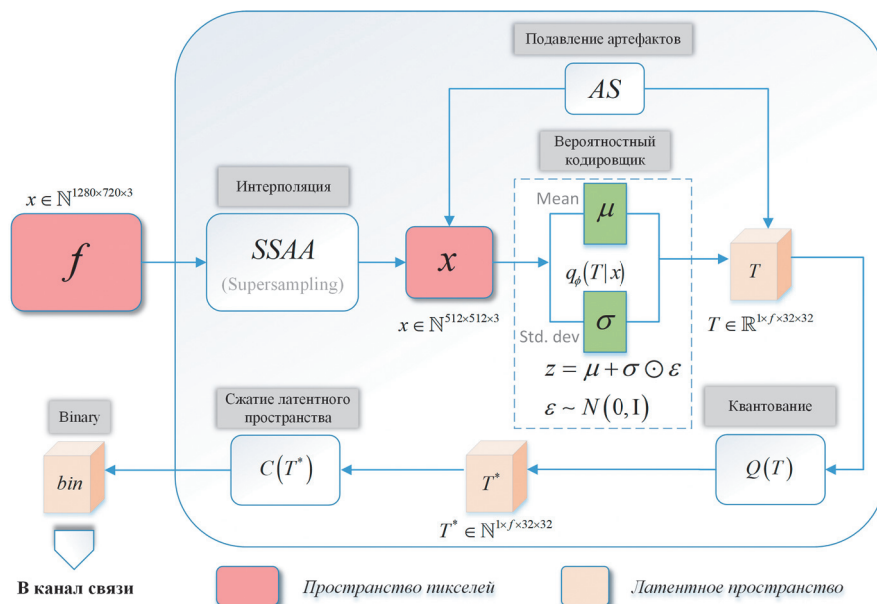
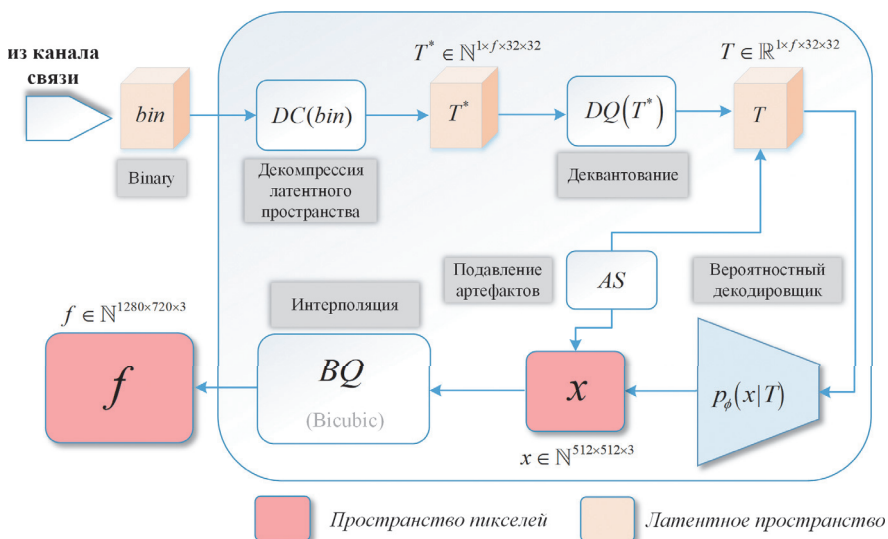


Рисунок 2
Модель нейросетевого декодера



различной длиной. Сетевые задержки и потери пакетов в сетях связи зависят от длины бинарной последовательности кадров. Следовательно, применение различных конфигураций позволит варьировать сетевые задержки и потери пакетов в зависимости от пропускной способности канала связи.

НЕЙРОСЕТЕВОЙ КОДЕК

В настоящей работе используется модификация нейросетевого кодера из исследований [1–3], который является развитием нейросетевого кодера, предложенного в [15, 16]. Модификация заключается в адаптации кодера для использования различных вариационных автокодировщи-

Рисунок 3

Преобразование размерности тензора для сжатия с потерями

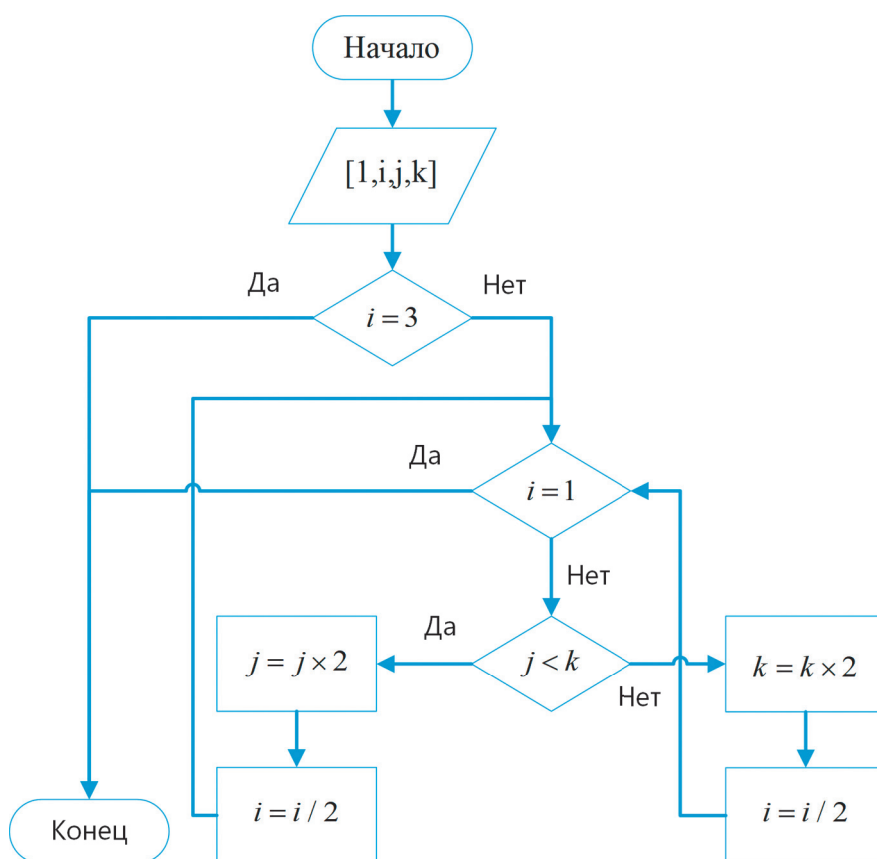
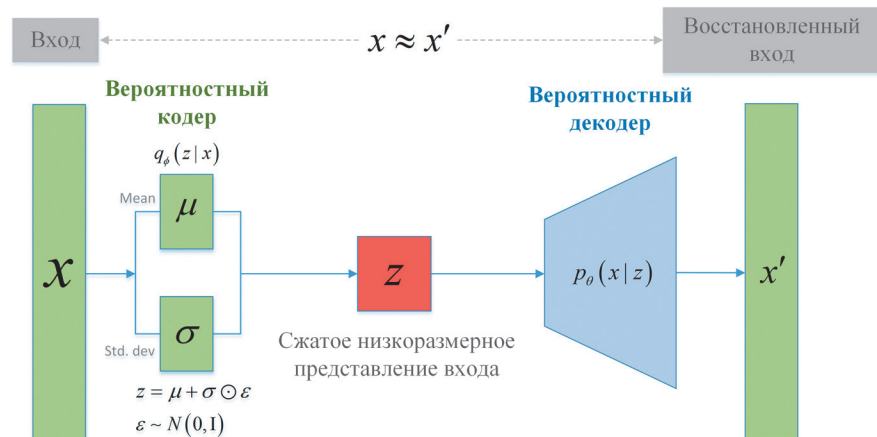


Рисунок 4

Вариационный автокодировщик



ков, а именно VQ-f4, VQ-f8, VQ-f16, KL-f4, KL-f8, KL-f16 и KL-f32, в блоках кодирования (в нейросетевом кодере) и декодирования (нейросетевом декодере).

Модели нейросетевого кодера и нейросетевого декодера показаны соответственно на рис. 1 и 2, где использованы следующие обозначения: f – исходный кадр разрешения 1280×720 ; x – интерполированный (методом суперсэмплинга) кадр до разрешения 512×512 ; T – ЛПП типа

float16, размерность и распределение которого зависят от используемого вариационного автокодировщика; T^* – квантованное ЛПП типа uint8; bin – сжатая бинарная последовательность на выходе кодера; $SSAA$ – интерполяция методом суперсэмплинга; BQ – обозначение бикубической интерполяции; AS – блок подавления артефактов; ε – распределённая по закону $N(0,1)$ случайная величина, где N – нормальный закон распределения; μ – среднее значение прямого распределения вариационного автокодировщика; σ – среднее квадратическое отклонение прямого распределения вариационного автокодировщика; $q_\theta(T|x)$ – прямое распределение вариационного автокодировщика; $p_\theta(x|T)$ – обратное распределение вариационного автокодировщика; $Q(T)$ – функция квантования; $DC(bin)$ – функция деквантования; $DQ(T^*)$ – функция деквантования.

Модификации нейросетевого кодера заключаются не только во встраивании других моделей вариационных автокодировщиков, но и в адаптации блоков квантования/деквантования и сжатия/деквантования для работы с размерностью ЛПП, отличающегося от (1, 8, 32, 32) для кадров разрешения (1, 3, 512, 512) в VQ-f16.

В предложенных для перевода ЛПП из float16 в uint8 алгоритмах квантования и соответствующих им алгоритмах деквантования параметры рассчитываются исходя из значений тензора ЛПП [2, 19, 20]. В то же время, если квантование выполняется на БС, а деквантование – на стороне оператора, то при деквантовании измеренные значения параметров оказываются неизвестны. Соответственно для решения данной проблемы в [2, 3] предложено использовать средние значения данных параметров, вычисленные на конкретном наборе данных. Тем не менее значения, рассчитанные в [2, 3], получены для ЛПП вариационного автокодировщика VQ-f16 и, соот-

ветственно, актуальны только для ЛПП, обладающих схожими распределениями. Для остальных автокодировщиков видов VQ и KL в блоки квантования $Q(T)$ и деквантования $DQ(T^*)$ нейросетевого кодера встроены средние значения параметров, полученные в настоящей работе.

Блоки сжатия $C(T^*)$ и декомпрессии $DC(bin)$ существуют в двух видах: без потерь и с потерями. Блок сжатия с потерями зависит от размерности ЛПП, и в предыдущих исследованиях он осуществлял приведение тензора T^* размерности $(1, 8, 32, 32)$ к одноканальному изображению $(1, 128, 64)$, что не подходит для случая ЛПП других размерностей. Модификация блока сжатия $C(T^*)$ заключается в динамическом преобразовании размерности тензора либо к одноканальному, либо к трёхканальному изображению.

Пусть есть квантованный тензор ЛПП T^* размерности $(1, i, j, k)$. Если $i=3$, то размерность T^* далее не преобразуется и алгоритм сжатия с потерями применяется к тензору T^* как к трёхканальному изображению. Если $i \neq 3$, то пока $i \neq 1$, итеративно осуществляется умножение наименьшей из j и k размерности на 2 и деление i на 2 (рис. 3). Таким образом формируется одноканальное изображение с наиболее равномерным отношением сторон, к которому применяется сжатие с потерями. Модификация блока декомпрессии $DC(b)$ в свою очередь заключается в явном задании размерности восстанавливаемого тензора, получаемой из используемого вариационного автокодировщика.

ВАРИАЦИОННЫЕ АВТОКОДИРОВЩИКИ

Вариационные автокодировщики [26] являются развитием стандартных автокодировщиков [27] и состоят из двух частей: кодировщик в сжатое низкоразмерное представление входа и декодировщик в обратном направлении. В отличие от классических автокодировщиков, которые преобразуют входные данные к некоторому конкретному тензору, вариационные осуществляют перевод к распределению (рис. 4) [16].

Пусть известны настоящие значения параметров распределения θ^* . Тогда, чтобы получить значение $x^{(i)}$, требуется получить латентное представление значения $z^{(i)}$ (сжатое низкоразмерное представление входа) из априорного распределения вариационного автокодировщика $p_\theta(z)$ и сгенерировать $x^{(i)}$ из условного вероятностного распределения $p_\theta(x|z=z^{(i)})$. Тогда оптимальными являются такие значения параметров распределения θ^* , что максимизируется функция правдоподобия априорного распределения [28]:

$$\theta^* = \arg \max \left(\prod_{i=1}^n p_\theta(x^{(i)}) \right).$$

θ^* также выражается через логарифм функции правдоподобия:

$$\theta^* = \arg \max \left(\sum_{i=1}^n \log(p_\theta(x^{(i)})) \right).$$

Но вычислять $p_\theta(x^{(i)})$ таким образом затруднительно:

$$p_\theta(x^{(i)}) = \int p_\theta(x^{(i)}|z) p_\theta(z) dz.$$

Для упрощения вводится функция аппроксимации $q_\phi(z|x)$ — вероятностный кодировщик, в то время как $p_\theta(x|z)$ — вероятностный декодировщик.

При обучении вариационных автокодировщиков необходимо учитывать, что оцениваемая функция аппроксимации $q_\phi(z|x)$ должна быть близка к настоящему апостериорному распределению $p_\theta(z|x)$. Их близость оценивается с помощью расстояния Кульбака—Лейблера [16, 29].

$$\begin{aligned} D_{KL}(q_\phi(z|x)|p_\theta(z|x)) &= \\ &= \int q_\phi(z|x) \log \left(\frac{q_\phi(z|x)}{p_\theta(z|x)} \right) dz = \\ &= \int q_\phi(z|x) \log \left(\frac{q_\phi(z|x) p_\theta(x)}{p_\theta(z, x)} \right) dz = \\ &= \int q_\phi(z|x) \left(\log(p_\theta(x)) + \log \left(\frac{q_\phi(z|x)}{p_\theta(z, x)} \right) \right) dz = \\ &= \log(p_\theta(x)) + \int q_\phi(z|x) \log \left(\frac{q_\phi(z|x)}{p_\theta(z, x)} \right) dz = \\ &= \log(p_\theta(x)) + \int q_\phi(z|x) \log \left(\frac{q_\phi(z|x)}{p_\theta(z|x) p_\theta(z)} \right) dz = \\ &= \log(p_\theta(x)) + E_{z \sim q_\phi(z|x)} \left(\log \left(\frac{q_\phi(z|x)}{p_\theta(z)} \right) - \log(p_\theta(z|x)) \right) = \\ &= \log(p_\theta(x)) + D_{KL}(q_\phi(z|x)|p_\theta(z)) - \\ &- E_{z \sim q_\phi(z|x)} (\log(p_\theta(z|x))). \end{aligned}$$

Таким образом, функцию потерь можно определить как

$$\begin{aligned} L_{VAE}(\theta, \phi) &= -\log p_\theta(z) + D_{KL}(q_\phi(z|x)|p_\theta(z|x)) = \\ &= D_{KL}(q_\phi(z|x)|p_\theta(z)) - E_{z \sim q_\phi(z|x)} (\log(p_\theta(z|x))). \end{aligned}$$

НАБОР ДАННЫХ

В настоящем исследовании используется два авторских набора данных [1, 2] (рис. 5), которые основаны на кадрах, сделанных во время полёта БС над полигоном, и извлечённых из видео согласно методике, представляющей собой модификацию методики [1, 2].

Рисунок 5

Пример кадров авторского набора данных



Первый набор является стандартным. Он состоит из 1000 кадров разрешения 1280×720 и предназначен для измерения метрик при определённых конфигурациях нейросетевого кодека. Второй набор данных является сокращённым. Он содержит 100 кадров разрешения 1280×720 и предназначен для проведения большого количества экспериментов и получения оценок метрик.

МЕТОДИКА ЭКСПЕРИМЕНТА

Достижение целей исследования требует выполнения следующих задач:

1. Исследование размерностей латентных пространств и времени выполнения кодирования и декодирования вариационными автокодировщиками.

2. Получение средних значений параметров сдвигов и масштабирования алгоритмов квантования для всех рассматриваемых вариационных автокодировщиков.

3. Исследование закономерностей в применении рассмотренных в [1–3] методов обработки ЛПП между вариационными автокодировщиками.

4. Получение лучших методов обработки ЛПП для каждого рассматриваемого вариационного автокодировщика и конфигурации блока сжатия кадров видеопотока нейросетевого кодека (под конфигурацией блока сжатия кадров понимается сочетание метода обработки ЛПП (алгоритмов сжатия и квантования) и вариационного автокодировщика).

5. Проверка гипотезы о том, что лучшие методы сжатия кадров видеопотока для некоторых режимов могут в зависимости от требований достигаться при

применении вариационных автокодировщиков с различным числом значений ЛПП.

Структура эксперимента состоит из обработки всех изображений набора данных нейросетевым кодеком определённой конфигурации: последовательно. Таким образом имитируется работа нейросетевого кодера при передаче кадров видеопотока от БС к оператору по линии связи с землёй.

В зависимости от набора данных было выделено два вида экспериментов: стандартный эксперимент (при использовании стандартного набора) и сокращённый эксперимент (при использовании сокращённого набора). Как было показано в [3], использование сокращённого набора позволяет измерять метрики качества восстанавливаемого кадра без статистически значимых различий (по t -тесту), но степень сжатия ЛПП и вероятность удовлетворения требований к качеству кадра P близки по абсолютным значениям, но обладают статистически значимыми различиями по t -тесту. P — это вероятность того, что для восстанавливаемого кадра видеопотока выполняются условия: индекс структурного сходства ($SSIM \geq 0,7975$); среднеквадратическая ошибка $MSE \leq 72,96$; пиковое отношение сигнала к шуму ($peak\ signal\ to\ noise\ ratio$) $PSNR \geq 29,5$. Данные оценки получены экспериментальным путём как результат соотнесения разборчивости восстанавливаемых кадров видеопотока с точки зрения возможности FPV-управления БС и их метрик качества: $SSIM$ и MSE .

Таким образом, применение сокращённых экспериментов в случае измерения CR , $MSize$ (размер сжатого представления латентного пространства) и P ограничено и используется для приблизительной оценки данных показателей при выборе лучших методов обработки ЛПП. После отбора измеренные показатели уточняются при помощи стандартных экспериментов.

Каждой задаче соответствует собственный этап эксперимента.

1. В рамках первого этапа экспериментов проводится исследование: распределения средних значений и среднеквадратических отклонений значений ЛПП вариационных автокодировщиков; времени их выполнения; размерности ЛПП. Сжатие $C(T)/DC(T)$ и квантование $Q(T)/DQ(T^*)$ в данном случае не используются. Время выполнения исследуется на оборудовании с использованием графического ускорителя Nvidia A100 80Gb.

Измеряемые показатели: t_{cd} (время кодирования в мс); t_{pcd} (время декодирования); $latent$ (размерность ЛПП); $SSIM$ (индекс структурного сходства восстановленного изображения); MSE (средний квадрат ошибки восстановленного изображения); $PSNR$ (пиковое отношение сигнала к шуму, дБ); m (среднее

значение ЛПП); σ (среднеквадратическое отклонение значений ЛПП). Из значений размерности ЛПП $latent$ рассчитывается $LSize$ (размер ЛПП) как произведение его измерений (4 измерения и разные размерности для различных вариационных автокодировщиков).

Выполняется восстановление плотности значений ЛПП с помощью ядерной оценки плотности [30, 31] с Гауссовым ядром для моделей VQ и KL-f32 путем вычисления полосы пропускания по Сильверману [32, 33]:

$$h = \left(\frac{4\sigma^5}{3n} \right)^{\frac{1}{5}},$$

где h — пропускная способность; n — размер выборки. Для моделей KL-f4 и KL-f16 — $h=2$, KL-f8 — $h=3$.

Уровень значимости для доверительных интервалов равен 5%.

Итого на данном этапе проводится 7 (по количеству рассматриваемых вариационных автокодировщиков) стандартных экспериментов. Эксперименты над VQ-f16 также проводятся для проверки и уточнения результатов исследований [1–3, 19, 20] и на этом, и на следующих этапах.

2. На втором этапе проводятся стандартные эксперименты, в ходе которых измеряются параметры сдвига и масштабирования алгоритмов квантования [2, 19, 20]. Их средние значения затем встраиваются в нейросетевой кодек. На этом этапе рассматриваются следующие алгоритмы квантования: линейный (ЛИН), степенной (СТ), логистический (ЛОГ), модифицированный логистический (мЛОГ), нечётно-степенной (НС3 и НС5), гиперболическо-тангенциальный (ГТ), модифицированный гиперболическо-тангенциальный (мГТ), двойной логистический (2ЛОГ), модифицированный двойной логистический (м2ЛОГ) и гиперболическо-синусоидальный (ГС). Итоговое сжатие ЛПП при этом не используется. Кроме того, измеряются метрики качества восстанавливаемого кадра ($SSIM$, MSE , $PSNR$) и рассчитывается вероятность удовлетворения требований к качеству кадра P [1–3].

На данном этапе проводится $77=7 \times 11$ экспериментов, где 7 — количество рассматриваемых моделей вариационных автокодировщиков, 11 — количество рассматриваемых алгоритмов квантования.

3. В ходе третьего этапа выполняется проверка соотношения степеней сжатия ЛПП CR [1–3] (процент того, на сколько уменьшается размер ЛПП в результате применения алгоритма сжатия) и метрик качества восстановленного кадра ($SSIM$, MSE , $PSNR$) при использовании одних методов обработки ЛПП с различными вариационными автокодировщиками. Для этого повторяются стандартные эксперименты из исследований [1–3]: всех алгоритмов сжатия без

потерь (Deflated, LZMA, GZip, BZip2, ZSTD, Brotli, LZ4, LZ4F, LZ4H5, LZW, LZF, LZFSE, AEC; WebP и JPEG LS в варианте сжатия без потерь с алгоритмом квантования ЛИН и алгоритмом сжатия без потерь LZMA со всеми алгоритмами квантования).

Измеряемые показатели: *SSIM*, *MSE*, *PSNR* и *MSize*. Расчётные показатели: *CR* и *P*. Всего на данном этапе проводится $182 = 7 \times (15 + 11)$ эксперимента, где 7 — количество рассматриваемых моделей вариационных автокодировщиков, 11 — количество рассматриваемых алгоритмов квантования, 15 — количество рассматриваемых алгоритмов сжатия без потерь.

4. На четвёртом этапе в рамках поиска лучших вариантов сжатия и квантования проводятся сокращённые эксперименты над алгоритмами сжатия и квантования, не показавшими себя неэффективными в предшествующих исследованиях [1–3]. На этом этапе исследуются только алгоритмы сжатия с потерями (AVIF, HEIC, JPEG, WebP, JPEG LS, JPEG XR и JPEG XL), так как алгоритмы сжатия без потерь показывают худшие результаты. Параметр *quality* принимает значение от 5 до 100 включительно с шагом 5; параметр *level* — от 1 до 9 включительно с шагом 1. На этом этапе рассматриваются алгоритмы квантования ЛИН и СТ и алгоритмы предварительного квантования BitRound и Scale с *nsd*=0 и 1. Далее над 100 лучшими методами обработки ЛПП для каждого вариационного автокодировщика проводятся стандартные эксперименты для уточнения полученных результатов.

Измеряемые показатели: *CR*, *SSIM* [34–37], *MSE* [37, 38], *PSNR* [36, 37], *MSize* и *P*. Для сравнения методов обработки ЛПП по методике, представленной в [1–3], сравниваются значения *pre_rank* (общая оценка соотношения степени сжатия и качества восстанавливаемых кадров).

При глобальном поиске лучших методов сжатия кадров видеопотока выполняется пересчёт CR_i (процент сжатия ЛПП относительно наибольшего размера ЛПП среди всех рассматриваемых моделей) при принятии размера ЛПП *LSize* равным максимальному среди всех рассматриваемых вариационных автокодировщиков, а именно для KL-f4 и VQ-f4 (1, 3, 128, 138), т.е. $1 \times 3 \times 128 \times 128 = 49152$ значений. В соответствии с изменениями *CR* выполняется пересчёт

pre_rank. Данные значения называются глобальным *pre_rank_r*.

Таким образом, на данном этапе проводится $10500 = 9800 + 700$ экспериментов, где $9800 = (2 \times 2 + 1) \times 2 \times 20 \times 7 \times 7$ сокращённых, где 2 — алгоритмы предварительного квантования с 2 значениями *nsd* (0 и 1) и без предварительного квантования, 2 — рассматриваемые алгоритмы квантования, 7 — алгоритмы сжатия с потерями с 20 значениями *quality*, 7 — количество рассматриваемых моделей вариационных автокодировщиков, и $700 = 7 \times 100$ стандартных, где 7 — количество рассматриваемых моделей вариационных автокодировщиков, 100 — количество экспериментов на каждую модель рассматриваемого вариационного автокодировщика.

5. На пятом этапе выполняется поиск лучших методов сжатия кадров видеопотока при $MSize_{cp} < MSize_{rp}$, где $MSize_{rp}$ — ограничение сверху по среднему минимальному размеру сжатого представления кадров видеопотока *bin*, $MSize_{cp}$ — средний размер сжатого представления латентного пространства. $MSize_{rp}$ принимает значения от 500 до 30000 байт включительно с шагом 500 байт. На этом этапе проверяется гипотеза о том, что на определённых отрезках $MSize_{rp}$ наилучшие методы сжатия кадров видеопотока могут обеспечивать разные вариационные автокодировщики. Сравнение методов сжатия кадров видеопотока осуществляется по *pre_rank_r*, полученном на четвёртом этапе.

Итого на всех этапах проводится 11466 экспериментов: 10500 сокращённых и 966 стандартных.

Во второй части статьи будет представлено исследование вариационных автокодировщиков и проведена оценка параметров алгоритмов квантования, будут исследованы закономерности методов обработки ЛПП, получены лучшие методы обработки ЛПП и предварительные конфигурации адаптивного нейросетевого кодера сжатия видеопотока. Также будут приведены результаты экспериментов.

Научная статья подготовлена в рамках прикладных научных исследований СПбГУТ, регистрационный номер 1023031600087-9-2.2.4; 2.2.5; 2.2.6; 1.2.1; 2.2.3 в ЕГИСУ НИОКТР

СПИСОК ЛИТЕРАТУРЫ

1. Березкин, А.А. Исследование методов латентного сжатия видеопотока при FPV управлении беспилотными системами / А.А. Березкин, А.А. Ченский, Р.В. Киричек, А.А. Захаров // Электросвязь. – 2024. – № 6. – С. 26–36.
2. Березкин, А.А. Исследование методов квантования латентного пространства вариационного автокодировщика для кадров FPV видеопотока. Часть I / А.А. Березкин, А.А. Ченский, Р.В. Киричек, А.А. Захаров // Электросвязь. – 2024. – № 6. – С. 10–16.
3. Березкин, А.А. Исследование методов квантования латентного пространства вариационного автокодировщика для кадров FPV видеопотока. Часть II / А.А. Березкин, А.А. Ченский, Р.М. Вивчарь, Р.В. Киричек // Электросвязь. – 2024. – № 7. – С. 26–35.
4. Стратегия развития отрасли связи Российской Федерации на период до 2035 года. – <http://government.ru/news/50304/> (дата обращения: 06.09.2024).
5. Xu, Q. Synthetic aperture radar image compression based on a variational

- p>autoencoder / Q. Xu, Y. Xiang, Z. Di et al. // IEEE Geoscience and Remote Sensing Letters. – 2021. – Vol. 19. – P. 1-5.
- 6. Chamain, L.D.** End-to-End Image Classification and Compression with variational autoencoders / L.D. Chamain, S. Qi, Z. Ding // IEEE Internet of Things Journal. – 2022. – Vol. 9, Issue 21. – P. 21916-21931.
- 7. Zhou, L.** Variational autoencoder for low bit-rate image compression / L. Zhou, C. Cai, Y. Gao et al. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. – 2018. – P. 2617-2620.
- 8. Yilmaz, M.A.** Self-organized variational autoencoders (self-vae) for learned image compression / M.A. Yilmaz, O. Keleş, H. Güven et al. // 2021 IEEE International Conference on Image Processing (ICIP). – IEEE, 2021. – P. 3732-3736.
- 9. Alves de Oliveira, V.** Reduced-complexity end-to-end variational autoencoder for on board satellite image compression / V. Alves de Oliveira, M. Chabert, T. Oberlin et al. // Remote Sensing. – 2021. – Vol. 13, Issue 3. – P. 447.
- 10. Sun, Y.** Image compression algorithm based on variational autoencoder / Y. Sun, L. Li, Y. Ding et al. // Journal of Physics: Conference Series. – 2021. – Vol. 2066, Issue 1. – P. 012008.
- 11. Wen, S.** Variational Autoencoder based Image Compression with Pyramidal Features and Context Entropy Model / S. Wen, J. Zhou, A. Nakagawa et al. // CVPR Workshops. – 2019. – P. 4321-4324.
- 12. Luo, J.** Noise-to-Compression Variational Autoencoder for Efficient End-to-End Optimized Image Coding / J. Luo, S. Li, W. Dai et al. // 2020 Data Compression Conference (DCC). – IEEE, 2020. – P. 33-42.
- 13. Liu, X.** Medical image compression based on variational autoencoder / X. Liu, L. Zhang, Z. Guo et al. // Mathematical Problems in Engineering. – 2022. – Vol. 2022, Issue 3. – P. 1-12.
- 14. Ballé, J.** Variational image compression with a scale hyperprior / J. Ballé, D. Minnen, S. Singh et al. // arXiv preprint arXiv:1802.01436. – 2018.
- 15. Березкин, А.А.** Метод сжатия видеопотока при управлении беспилотными системами в гибридных орбитально-наземных сетях связи / А.А. Березкин, Р.М. Вивчарь, А.В. Слепнев и др. // Электросвязь. – 2023. – № 10. – С. 48-56.
- 16. Березкин, А.А.** Метод декомпрессии FPV-видеопотока от беспилотных систем на основе латентной диффузионной нейросетевой модели / А.А. Березкин, Р.М. Вивчарь, Р.В. Киричек, А.А. Захаров // Электросвязь. – 2024. – № 1. – С. 25-36.
- 17. Ronneberger, O.** U-net: Convolutional networks for biomedical image segmentation / O. Ronneberger, P. Fischer, T. Brox // 18th international conference Medical image computing and computer-assisted intervention (MICCAI 2015), proceedings, part III. – Springer International Publishing, 2015. – P. 234-241.
- 18. Березкин, А.А.** Многокритериальная оценка эффективности управления беспилотными системами в гибридных сетях связи / А.А. Березкин, Р.М. Вивчарь, Р.В. Киричек // Труды учебных заведений связи. – 2024. – Т. 10, № 1. – С. 18-25.
- 19. Ченский, А.А.** Исследование методов квантования при сжатии видеопотока для управления беспилотными системами от первого лица / А.А. Ченский, А.А. Березкин // Актуальные проблемы инфотелекоммуникаций в науке и образовании 2024 (АПИНО-2024): международная конференция. – СПб., 2024. – С. 107-111.
- 20. Ченский, А.А.** Квантование видеопотока при FPV-управлении беспилотными системами в гибридных сетях связи / А.А. Ченский, А.А. Березкин // 79-я научно-техническая конференция СПб НТО РЭС им. А.С. Попова, посвященная Дню радио. Сборник докладов. – СПб., 2024. – С. 186-189.
- 21. Березкин, А.А.** Исследование алгоритмов сжатия данных как элемента системы информационного обмена для управления беспилотными комплексами в гибридных сетях связи / А.А. Березкин, Р.М. Вивчарь, А.А. Лазарев // Всероссийская научно-техническая и научно-методическая конференция магистрантов и их руководителей (ПКМ-2023); Сборник лучших докладов. – СПб.: СПбГУТ, 2023. – С. 177-182.
- 22. Зуйков М. А.** Анализ эффективности алгоритмов сжатия данных в каналах информационного обмена БАС в гибридных орбитально-наземных сетях связи / М.А. Зуйков, А.А. Березкин, Р.В. Вивчарь, А.В. Пачин // 79-я научно-техническая конференция СПб НТО РЭС им. А.С. Попова. Сборник докладов. – СПб., 2024. – С. 197-199.
- 23. Stable Diffusion: Model Zoo.** – <https://github.com/pesser/stable-diffusion/tree/main#model-zoo/> (дата обращения: 06.09.2024).
- 24. Rombach, R.** High-resolution image synthesis with latent diffusion models / R. Rombach, A. Blattmann, D. Lorenz et al. // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2022. – P. 10684-10695.
- 25. Березкин, А.А.** Адаптивный контроль интенсивности видеопотока при передаче FPV-трафика беспилотных систем / А.А. Березкин, А.А. Паршин, А.А. Лазарев // 79-я научно-техническая конференция СПб НТО РЭС им. А.С. Попова. Сборник докладов. – СПб., 2024. – С. 162-165.
- 26. Kingma, D.P.** Auto-encoding variational bayes / D.P. Kingma, M. Welling // arXiv preprint arXiv:1312.6114.
- 27. Hinton, G. E.** Reducing the dimensionality of data with neural networks / G.E. Hinton, R.R. Salakhutdinov // Science. – 2006. – Vol. 313, № 5786. – P. 504-507.
- 28. Weng, L.** From Autoencoder to Beta-VAE / L. Weng. – <https://lilianweng.github.io/posts/2018-08-12-vae/#vae-variational-autoencoder> (дата обращения: 30.07.2024).
- 29. Kullback S.** On information and sufficiency / S. Kullback, R.A. Leibler // The Annals of Mathematical Statistics. – 1951. – Vol. 22, Issue 1. – P. 79-86.
- 30. Rosenblatt, M.** Remarks on Some Nonparametric Estimates of a Density Function / M. Rosenblatt // The Annals of Mathematical Statistics. – 1956. – Vol. 27, Issue 3. – P. 832-837.
- 31. Parzen, E.** On Estimation of a Probability Density Function and Mode / E. Parzen // The Annals of Mathematical Statistics. – 1962. – Vol. 33, Issue 3. – P. 1065-1076.
- 32. Silverman, B. W.** Density estimation for statistics and data analysis / B.W. Silverman. – Routledge, 2018. – 176 p.
- 33. Chiu, S.T.** Bandwidth selection for kernel density estimation / S.T. Chiu // The Annals of Statistics. – 1991. – Vol. 19, Issue 3. – P. 1883-1905.
- 34. Bakurov, I.** Structural similarity index (SSIM) revisited: A data-driven approach / I. Bakurov, M. Buzzelli, R. Schettini et al. // Expert Systems with Applications. – 2022. – Vol. 189. – article 116087.
- 35. Nilsson, J.** Understanding SSIM / J. Nilsson, T. Akenine-Möller. – arXiv preprint arXiv:2006.13846.
- 36. Hore, A.** Image quality metrics: PSNR vs. SSIM / A. Hore, D. Ziou // 2010

20th international conference on pattern recognition. – IEEE, 2010. – P. 2366-2369.

37. Sara, U. Image quality assessment through FSIM, SSIM, MSE and PSNR – a

comparative study / U. Sara, M. Akter, M.S. Uddin // Journal of Computer and Communications. – 2019. – Vol. 7, Issue 3. – P. 8-18.

38. Schluchter, M.D. Mean square error / M.D. Schluchter // Encyclopedia of Biostatistics. – John Wiley & Sons, 2005.

Получено 09.09.24

Research of neural network codec configurations for adaptive FPV video stream frame compression system when controlling unmanned systems. Part I: Methodology

A.A. Berezkin, Bonch-Bruевич State university of telecommunications (SPbSUT), associate professor, Ph.D.; berezkin.aa@sut.ru

A.A. Chenskiy, SPbSUT, masters student; chenskii.aa@sut.ru

R.V. Kirichek, SPbSUT, rector, professor, Dr.Sci.(Tech.); kirichek@sut.ru

A.A. Zaharov, The M.I. Krivosheev Radio Research & Development Institute (NIIR), deputy director general, associate professor, Ph.D.; a.zaharov@niir.ru

Abstract. In this work we investigate different configurations of neural network variational autoencoders of VQ and KL types (VQ-f4, VQ-f8, VQ-f16, KL-f4, KL-f8, KL-f16, KL-f32) from the Stable Diffusion models for the possibility of their use for video stream frame compression in FPV control of unmanned systems. The first part of the paper summarizes the results of studies of compression and quantization of latent spaces of variational autoencoders, presents a modification of neural network codec, data sets and discusses the methodology of experiments.

Keywords: variational autoencoder, quantization, data compression, neural networks, latent space, video stream compression, FPV control, UAV.