
**РАСПОЗНАВАНИЕ ОБРАЗОВ
И ОБРАБОТКА ИЗОБРАЖЕНИЙ**

УДК 004.932.2

**КОМПЛЕКСИРОВАНИЕ ИЗОБРАЖЕНИЙ РАЗНЫХ
СПЕКТРАЛЬНЫХ ДИАПАЗОНОВ НА ОСНОВЕ
ГЕНЕРАТИВНЫХ СОСТЯЗАТЕЛЬНЫХ СЕТЕЙ¹**

© 2019 г. Ю. В. Визильтер^а, О. В. Выголов^а, Д. В. Комаров^а, М. А. Лебедев^{а,*}

^аФГУП ГосНИИАС, Москва, Россия

*e-mail: MLebedev@gosniias.ru

Поступила в редакцию 03.12.2018 г.

После доработки 25.12.2018 г.

Принята к публикации 28.01.2019 г.

Предлагается метод комплексирования разносектральных изображений с использованием генеративных состязательных сетей. Разработана оригинальная архитектура сети “FusionNet”, основанная на архитектуре pix2pix, которая обеспечивает построение комплексного (интегрального) изображения, объединяющего наиболее информативные фрагменты разносектральных изображений и, как следствие, обладающего большей информативностью по сравнению с каждым из этих изображений в отдельности. Приводятся описания методики построения обучающей и тестовой выборок, а также процесса аугментации данных. Показаны примеры работы предлагаемого метода комплексирования на реальных телевизионных и инфракрасных изображениях.

DOI: 10.1134/S000233881903020X

Введение. Проблема создания эффективных методов и алгоритмов комплексирования изображений, регистрируемых датчиками технического зрения в различных спектральных диапазонах, по-прежнему актуальна как в задачах обеспечения надежного автоматического обнаружения и распознавания объектов в широком диапазоне условий их наблюдения, так и в задачах человеко-машинного взаимодействия в части информационной поддержки оператора средствами улучшенного видения операционной среды [1]. В частности, в авиационных системах улучшенного видения (СУВ) для обеспечения информационной поддержки экипажа в широком диапазоне условий видимости актуальна задача комплексирования информации, получаемой от разносектральных датчиков технического зрения.

На сегодняшний день основной парадигмой в области обработки и анализа изображений являются глубокие конволюционные нейронные сети (ГКНС) и глубокое машинное обучение. Практически во всем спектре задач технического зрения нейросетевые алгоритмы на основе ГКНС демонстрируют качественно лучшие результаты по сравнению с алгоритмами предыдущих поколений, приближая эти результаты к возможностям человека, а в ряде случаев даже превосходя их. ГКНС могут эффективно применяться в том числе и при решении задач комплексирования разносектральных изображений.

В данной работе предлагается новый метод комплексирования разносектральных изображений на основе генеративной состязательной нейронной сети. Отличительными особенностями метода являются формирование комплексного изображения с низким уровнем шумов, сохранение информативных областей со всех входных разносектральных изображений, отсутствие артефактов, характерных для классических методов комплексирования, например, наличие ореолов свечения в случае метода комплексирования с использованием пирамиды Лапласианов, решение проблемы локальной инверсии контраста, устранение геометрического рассогласования входных изображений на комплексном изображении.

1. Генеративно-состязательные сети. Комплексирование разносектральных изображений осуществляется на основе некоторой признаковой информации об изображениях, т.е. подразумевает объединение в комплексном (интегральном) изображении наиболее информативных

¹ Работа выполнена при поддержке РФФ (грант № 16-11-00082) и РФФИ (грант № 18-07-01275А).

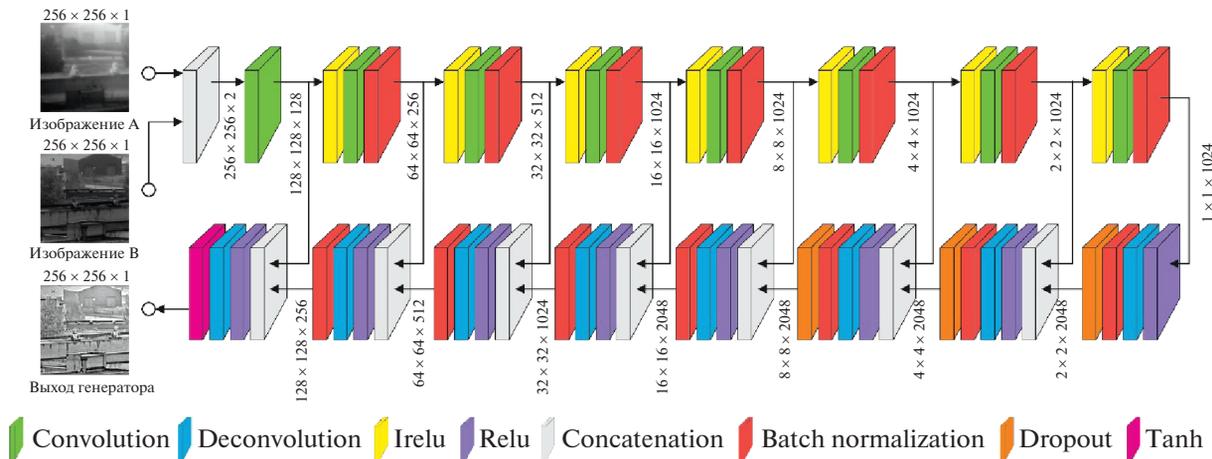


Рис. 1. Архитектура сети генератора

фрагментов с каждого из разномасштабных изображений. На сегодняшний день наилучшие результаты в решении проблемы сопоставления признаков достигаются с использованием особого вида нейронных сетей, так называемых генеративно-состязательных сетей (ГСС).

ГСС – алгоритм машинного обучения без учителя, построенный на комбинации из двух конкурирующих нейронных сетей: генератора G и дискриминатора D . Генератор G обучается создавать изображения y на основе некоторого пространства скрытых (латентных) признаков, извлеченных из входного изображения x , на котором имеется случайный шум z :

$$G = G(x, z) : \{x, z\} \rightarrow y.$$

Дискриминатор D в свою очередь обучается хорошо обнаруживать “поддельные” изображения генератора G :

$$D = D(x, y) : \{x, y\} \rightarrow [0, 1].$$

Дискриминатор отображает объекты из пространства данных в отрезок $[0, 1]$, который интерпретируется как вероятность того, что пример был действительно “настоящий”, а не синтезированный генератором G .

Фактически дискриминатор и генератор играют между собой в игру, которая в теории игр называется минимаксной игрой, решая следующую оптимизационную задачу:

$$GAN = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D).$$

общая функция потерь ГСС $\mathcal{L}_{GAN}(G, D)$ имеет вид

$$\mathcal{L}_{GAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))],$$

где $\mathbb{E}_{x,y}$, $\mathbb{E}_{x,z}$ – математическое ожидание по обучающему набору.

2. Архитектура сети FusionNet. Предлагаемая нами сеть FusionNet основывается на широко распространенной архитектуре pix2pix [2] и также состоит из двух основных частей: генератора и дискриминатора. Отличительная особенность нашей реализации заключается в том, что генератор обрабатывает одновременно все входные разномасштабные изображения, извлекая из каждого характерные признаки. Текущая реализация сети работает с двумя входными разномасштабными изображениями, но это количество может быть увеличено без существенного изменения архитектуры самой сети. К входным изображениям применяется процедура конкатенации по глубине, т.е. для входных n изображений формируется n -мерный тензор, а затем полученная таким образом “стопка” изображений подается на вход генератору. В основе генератора используется сеть “U-Net” [3]. Архитектура сети генератора приведена на рис. 1.

Дискриминатор – это другая отдельная сеть, архитектура которой называется “PatchGAN” [2]. В нашей реализации на вход ей поступает три изображения: два разномасштабных изображения и комплексированное изображение для них, причем в качестве комплексированного

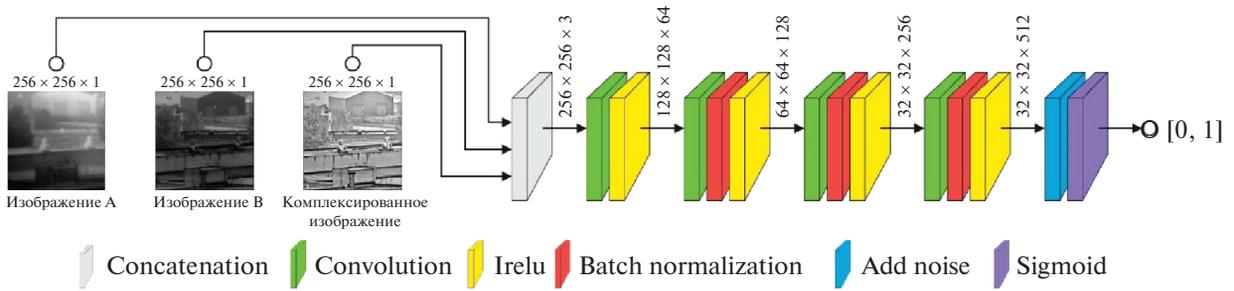


Рис. 2. Архитектура сети дискриминатора

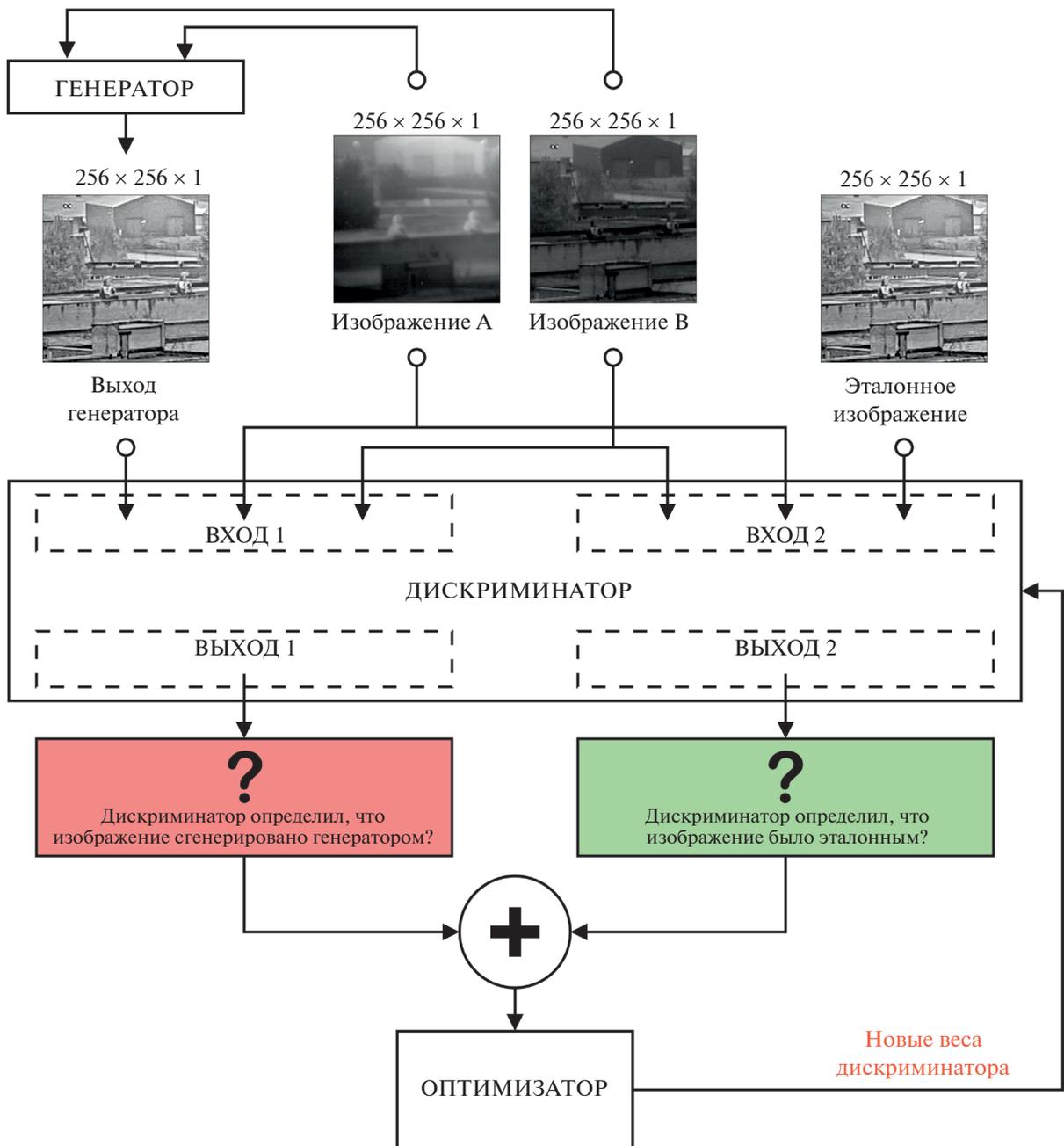


Рис. 3. Блок-схема обучения дискриминатора

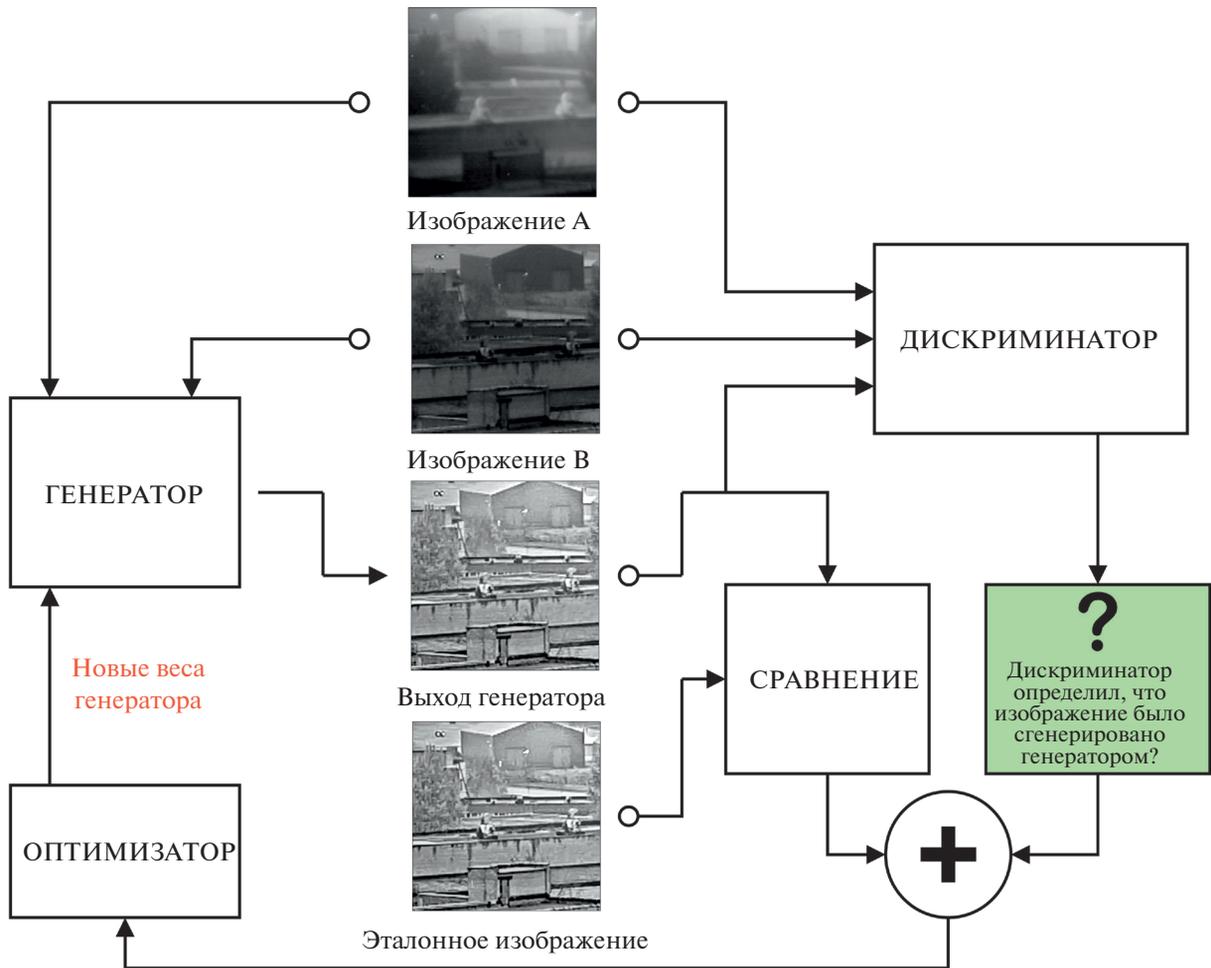


Рис. 4. Блок-схема обучения генератора

изображения в одном случае подается результат работы генератора, а в другом — эталонное комплексированное изображение. Суть работы дискриминатора заключается в том, чтобы научиться определять принадлежность комплексированного изображения к классу сгенерированных генератором или же к классу эталонных. Структура архитектуры дискриминатора очень похожа на кодирующую часть генератора, но выходом является одно число в диапазоне от 0 до 1, демонстрирующее меру правдоподобия комплексированного изображения и соответствующих входных разносектральных изображений. Архитектура сети дискриминатора приведена на рис. 2.

Процесс обучения сети состоит из двух шагов: обучение дискриминатора и обучение генератора.

Для обучения дискриминатора генератор генерирует комплексированное изображение, а затем дискриминатор оценивает входные разносектральные изображения с результатом работы генератора и с эталонным комплексированным изображением, после чего дает свое представление о том, насколько они реалистичны. После этого веса генератора корректируются на основе ошибки классификации. Блок-схема обучения дискриминатора представлена на рис. 3.

Затем корректируются веса генератора на основе выходного сигнала дискриминатора и разницы выходного комплексированного изображения, сгенерированного генератором, с эталонным комплексированным изображением. Блок-схема обучения генератора рассмотрена на рис. 4.

В обоих случаях в качестве оптимизатора используется метод стохастической оптимизации Adam [4]. Финальная функция потерь имеет вид как в [2]:

$$GAN = \arg \min_G \max_D [\mathcal{L}_{GAN}(G, D) + \lambda \mathcal{L}_L(G)],$$



Рис. 5. Пример входных разносектральных изображений: слева – ТВ-изображение, справа – ИК-изображение (3–5 мкм)

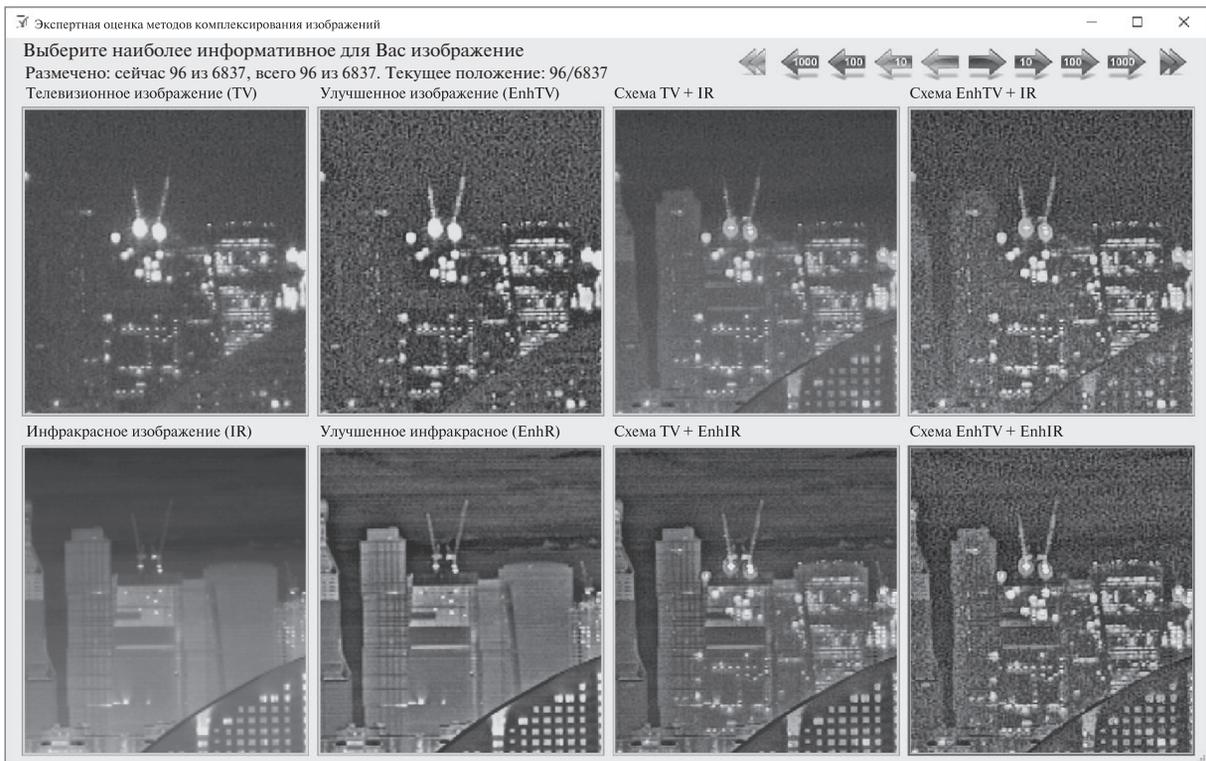


Рис. 6. Пример окна программы для экспертной оценки схем комплексирования изображений

$$\mathcal{L}_{L_1}(G) = \mathbb{E}_{x,y,z} [\|y - G(x, z)\|_1],$$

где λ – коэффициент, синхронизирующий значения $\mathcal{L}_{GAN}(G, D)$ и $\mathcal{L}_{L_1}(G)$;

$\mathcal{L}_{L_1}(G)$ – L_1 норма эталонного и сгенерированного генератором изображений;

$\mathbb{E}_{x,y,z}$ – математическое ожидание по обучающему набору.

3. Подготовка обучающей и тестовой выборки. Для формирования обучающей выборки помимо входных разносектральных изображений также необходимы соответствующие им эталонные комплексированные изображения. На сегодняшний день существует большое количество

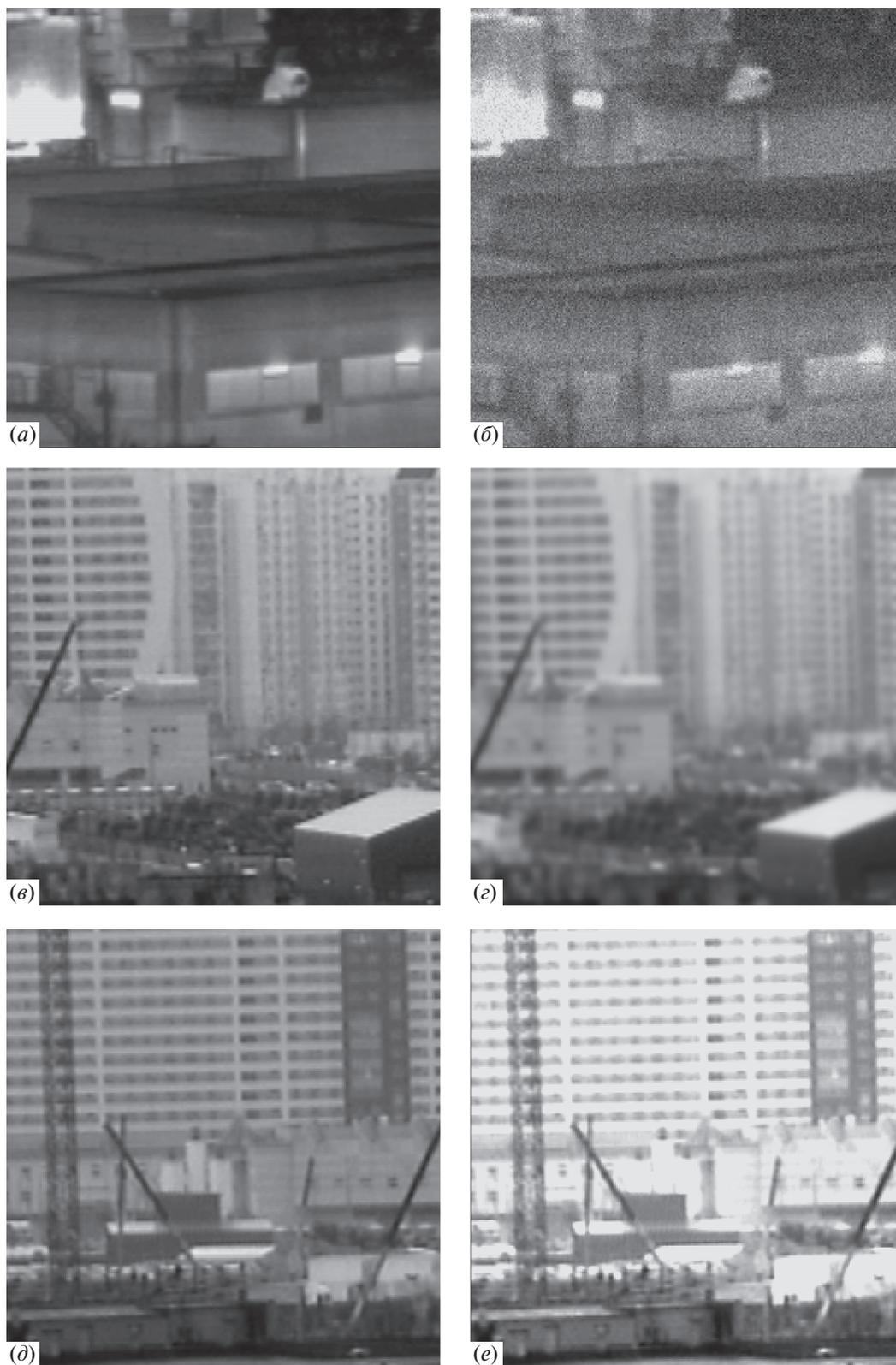


Рис. 7. Пример аугментации входных изображений: *a* – входное ИК-изображение, *б* – зашумленное изображение “*a*”, *z* – входное ТВ-изображение, *z* – размытое изображение “*z*”, *д* – входное ТВ-изображение, *е* – изображение “*д*” с измененной яркостью

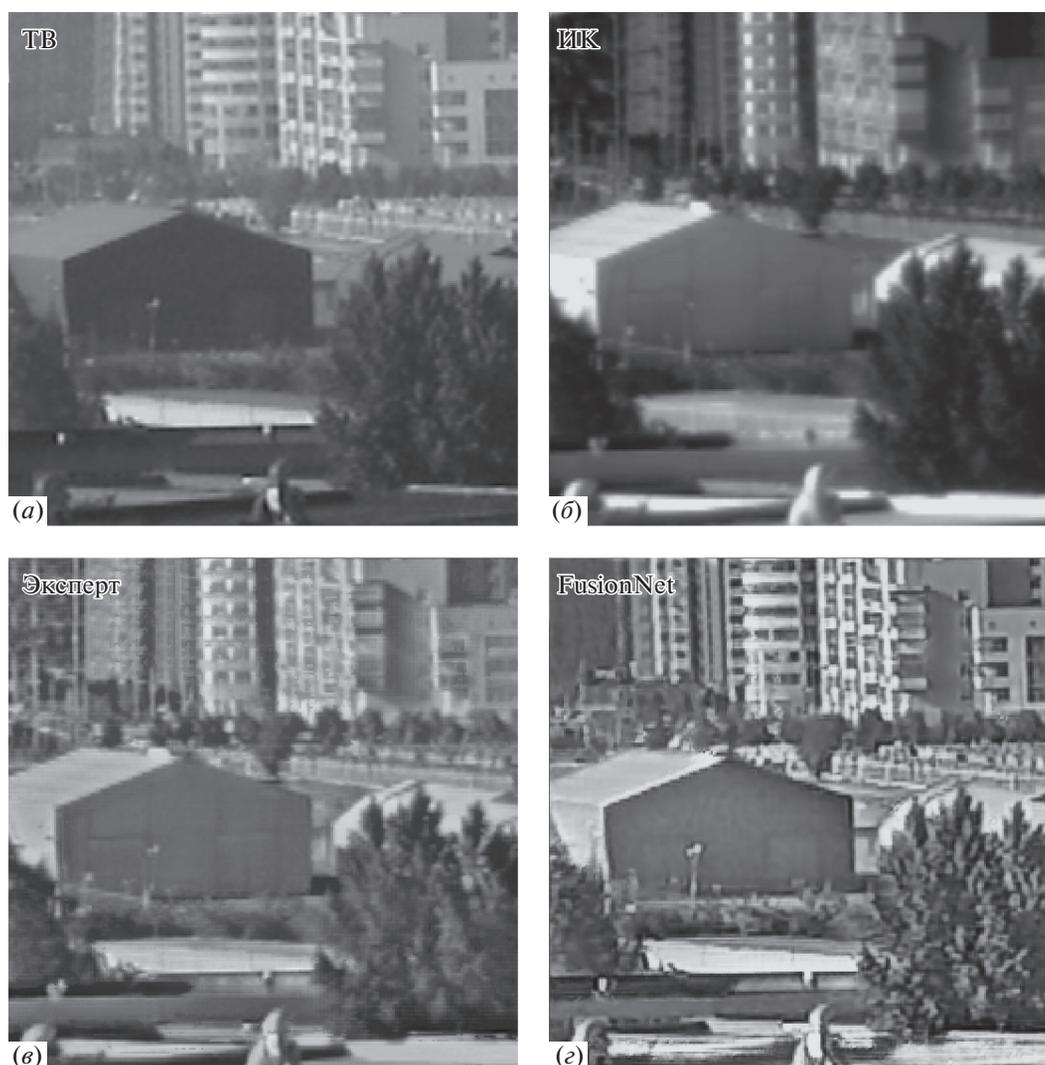


Рис. 8. Пример повышения резкости при комплексировании для изображений, полученных при хороших погодных условиях: *a* – входное ТВ-изображение, *b* – входное ИК-изображение, *в* – результат комплексирования, выбранный экспертом, *z* – результат комплексирования, выбранный сетью FusionNet

методов комплексирования разномасштабных изображений [1, 5–10], в том числе основанных на глубоких нейронных сетях [11]. В данной работе для формирования эталонных комплексированных изображений использовался метод, основанный на пирамиде Лапласианов [5].

В качестве входных разномасштабных изображений использовались изображения, регистрируемые в видимом телевизионном (ТВ) и инфракрасном (ИК) 3–5 мкм диапазонах. Изображения регистрировались оптическими каналами, не имеющими взаимной аппаратной юстировки.

Для обучения и тестирования было отобрано 25 видеороликов различных сцен, полученных при разных погодных условиях и в разное время суток. Совокупное отобранное количество изображений из видеороликов составило 6837. ТВ- и ИК-изображения юстировались программным методом, и итоговые изображения обрезались под формат 636×514 пикселей. Пример входных разномасштабных изображений приведен на рис. 5.

Результат комплексирования на основе Лапласиана зачастую можно улучшить, предварительно применив к какому-нибудь одному или сразу ко всем разномасштабным изображениям алгоритм улучшения MultiScale Retinex (MSR) [12]. При этом одному человеку может казаться более информативным результат комплексирования с применением алгоритма улучшения, например, к ТВ-изображению, а другому – к ТВ-, и к ИК-изображениям. В данной работе под

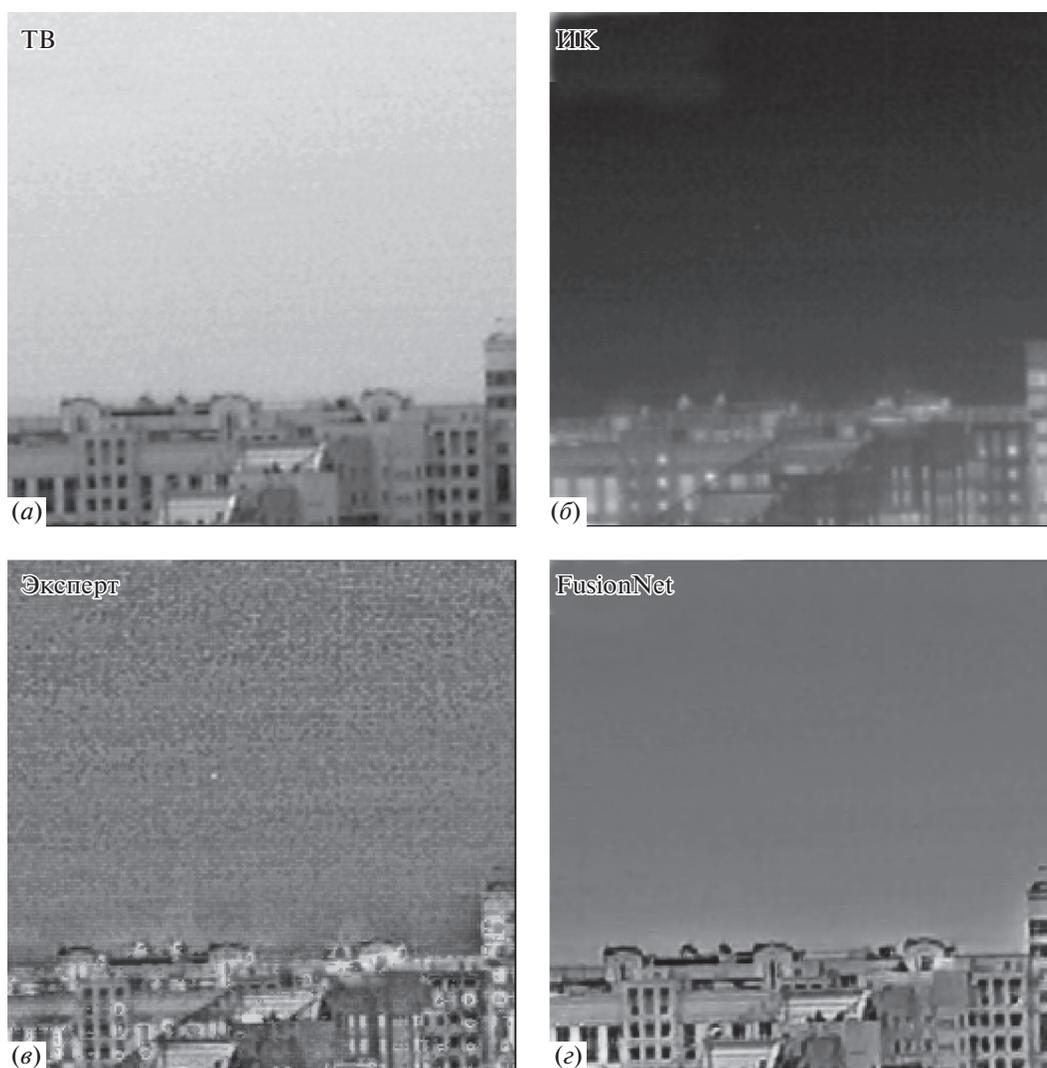


Рис. 9. Пример устранения шумов сетью FusionNet: *a* – входное ТВ-изображение, *б* – входное ИК-изображение, *в* – результат комплексирования, выбранный экспертом, *г* – результат комплексирования, выбранный сетью FusionNet

информативностью изображения понимается не столько художественная оценка качества изображения, сколько оценка максимального визуального восприятия количества объектов на изображении и их визуальных характеристик.

Для создания наиболее полной и независимой обучающей выборки использовалась экспертная оценка. Для этого было разработано специальное программное приложение, в котором авторизованным экспертом выбиралось наиболее информативное на его взгляд изображение. Эксперту предлагалось выбрать из следующих примеров:

- 1) ТВ-изображение;
- 2) ИК-изображение;
- 3) улучшенное ТВ-изображение, к которому применили алгоритм MultiScale Retinex (EnhTV);
- 4) улучшенное ИК-изображение, к которому применили алгоритм MultiScale Retinex (EnhIR);
- 5) комплексированное изображение исходных ТВ- и ИК-изображений (схема TV + IR);
- 6) комплексированное изображение исходного ТВ- и улучшенного ИК-изображений (схема TV + EnhIR);
- 7) комплексированное изображение улучшенного ТВ- и исходного ИК-изображений (схема EnhTV + IR);

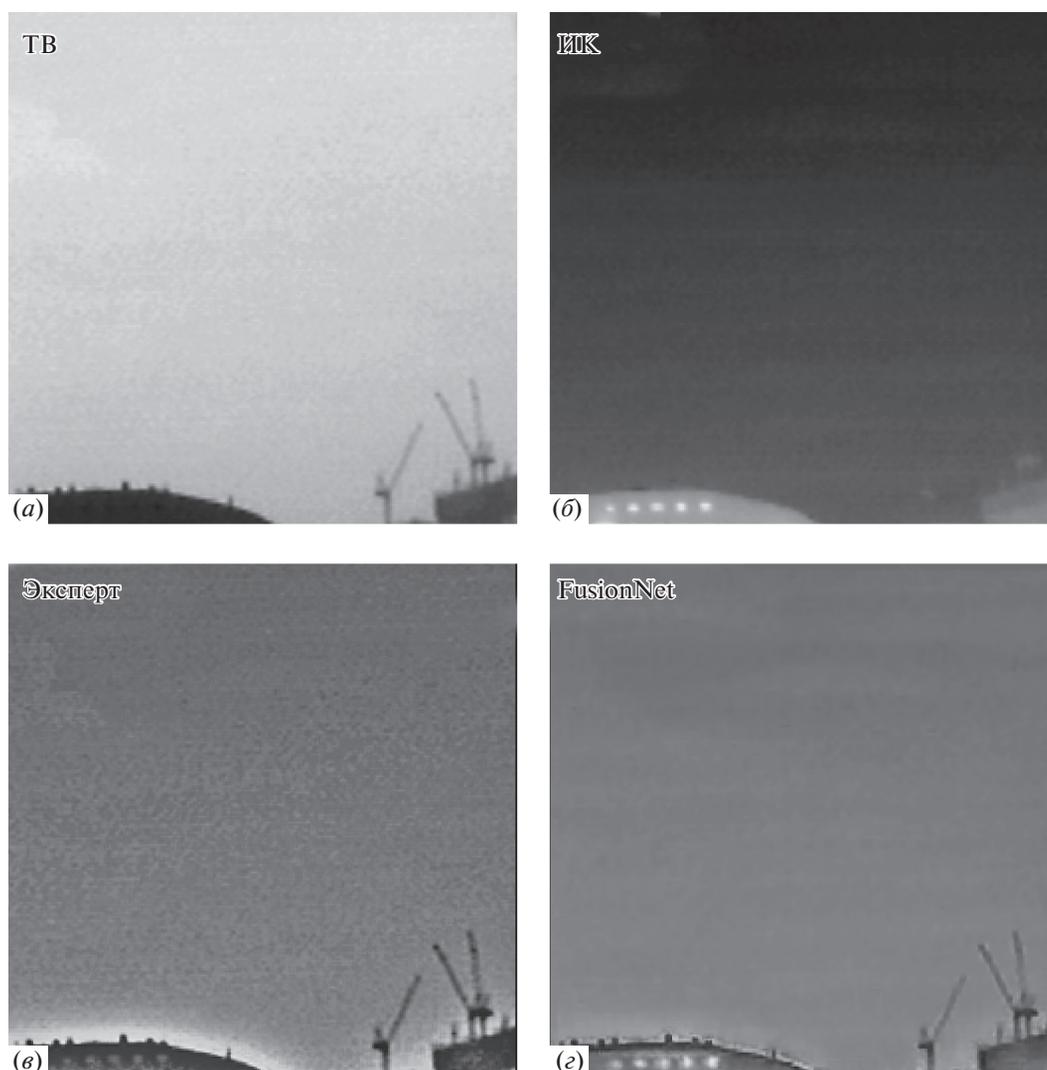


Рис. 10. Пример устранения артефактов в виде ореолов сетью FusionNet: *a* – входное ТВ-изображение, *б* – входное ИК-изображение, *в* – результат комплексирования, выбранный экспертом, *г* – результат комплексирования, выбранный сетью FusionNet

8) комплексированное изображение улучшенных ТВ- и ИК-изображений (схема EnhTV + EnhIR).

Пример интерфейса программного приложения для экспертной оценки схем комплексирования изображений представлен на рис. 6.

Для экспертной оценки привлекалось шесть экспертов, которые в совокупности разместили 19.5 тыс. изображений. Как видно из рис. 6, далеко не все изображения являются идеальными. Комплексированное изображение может содержать в себе шум, артефакты метода комплексирования в виде ореолов, размытие и т.д. И так как целью работы является создание наилучшего метода комплексирования, а не повторения результатов комплексирования рассмотренными схемами, то для обучающей выборки были отобраны 5453 изображения, полученные при хороших условиях съемки.

В качестве тестовой выборки были отобраны случайным образом 3 тыс. изображений из набора, полученного путем экспертной оценки, за исключением тех изображений, которые уже были отобраны в обучающую выборку. Изображения из тестовой выборки были получены при абсолютно разных условиях съемки.

4. Аугментация данных. Применение разработанной сети FusionNet предполагается при разных погодных условиях, в том числе и при плохих, поэтому перед началом обучения для имита-

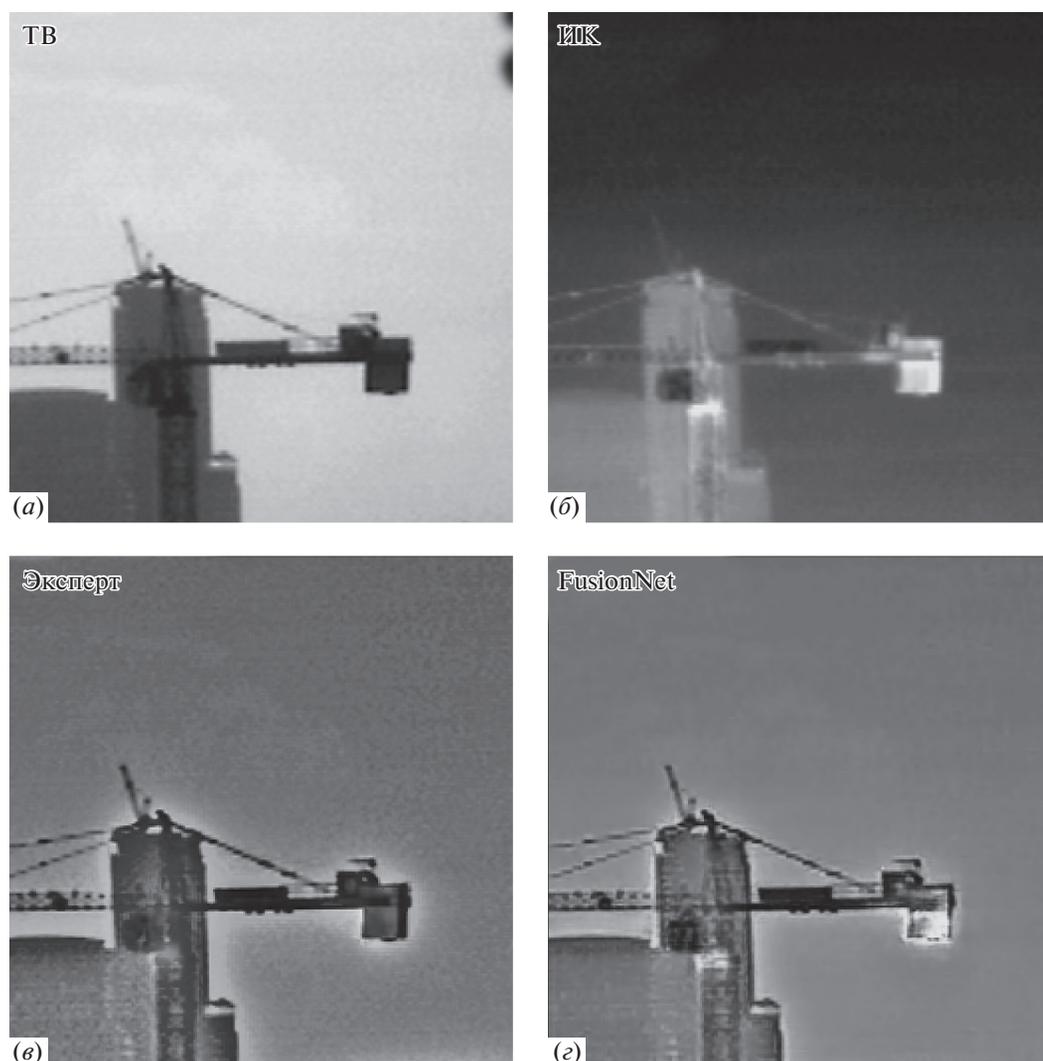


Рис. 11. Пример комплексирования изображений с сильной инверсией контраста на объекте: *a* – входное ТВ-изображение, *б* – входное ИК-изображение, *в* – результат комплексирования, выбранный экспертом, *г* – результат комплексирования, выбранный сетью FusionNet

ции плохих условий съемки добавляется слой аугментации данных, который должен обеспечить искажение входных изображений ТВ- и ИК-диапазонов. При этом эталонное комплексированное изображение не должно подвергаться искусственным искажениям, так как на выходе мы хотим иметь результат, близкий к результату, который был бы получен при идеальных условиях съемки.

Входные разномасштабные изображения имеют разрешения 636×514 пикселей, а предлагаемая архитектура сети работает с изображениями 256×256 пикселей. Поэтому на первом этапе аугментации данных происходит вырезание фрагментов размером 256×256 пикселей из входных изображений. При этом координаты вырезаемого фрагмента на разномасштабных изображениях задаются случайным образом и имеют случайное расхождение до 5 пикселей по осям абсцисс и ординат. Это было сделано для имитации рассогласования кадров, которое может возникнуть при резком перемещении системы технического зрения за счет отсутствия синхронизации камер. Далее, для повышения разнообразия, полученные фрагменты зеркально отражаются по вертикали и/или горизонтали.

Так как все изображения из обучающей выборки близки к идеальным по условиям съемки, то имитация времени суток осуществлялась простым случайным изменением параметров яркости для изображений. Имитация изображений не в фокусе или тумана для ТВ-канала осуществлялась добавлением случайного гауссовского размытия с максимальной сигмой $\sigma = 2$. Для имита-

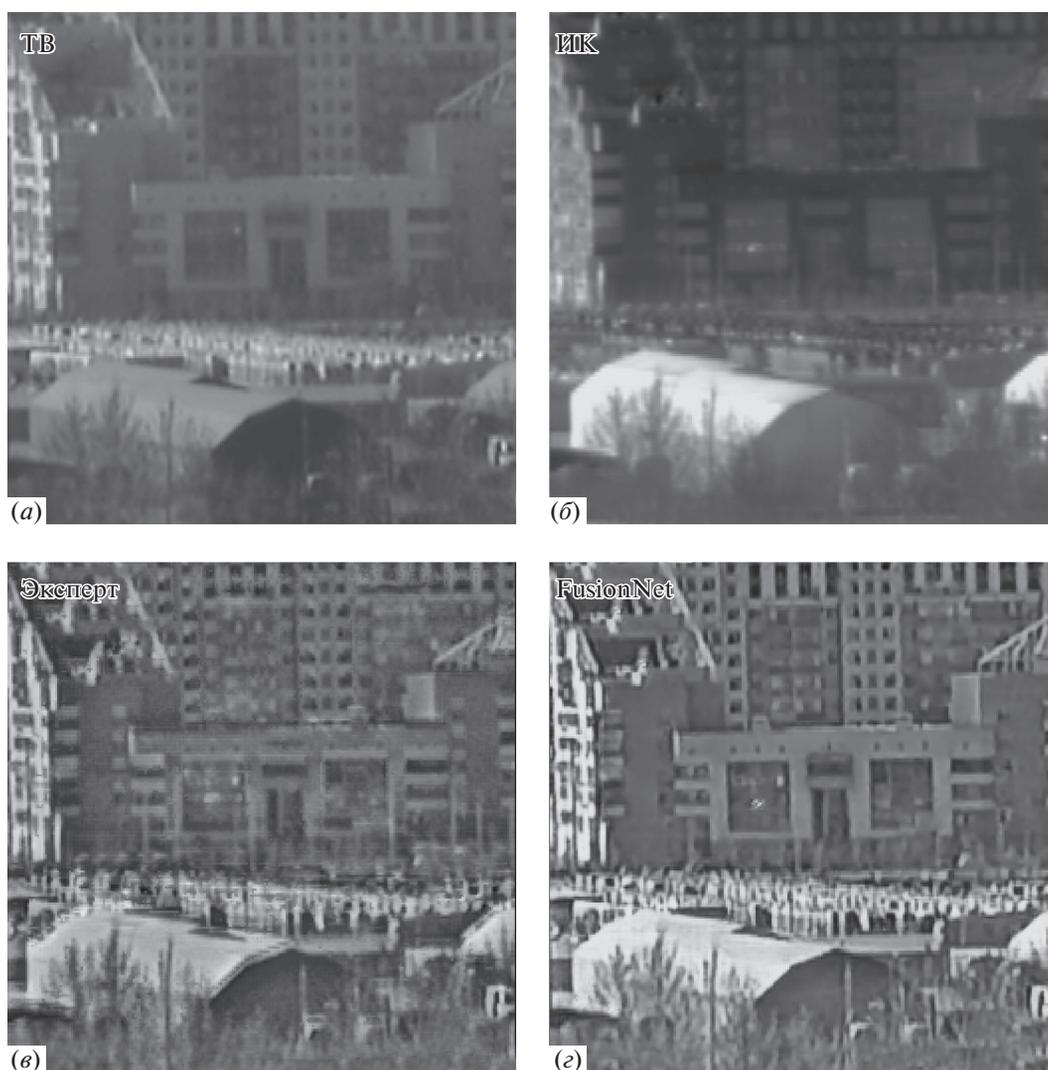


Рис. 12. Пример внутренней юстировки изображений сетью FusionNet: *a* – входное ТВ-изображение, *б* – входное ИК-изображение, *в* – результат комплексирования, выбранный экспертом, *г* – результат комплексирования, выбранный сетью FusionNet

ции характерных для реальных сенсоров помех, особенно в условиях недостаточной освещенности, на входные изображения добавлялся случайный аддитивный гауссовский шум с нулевым математическим ожиданием и максимальной дисперсией $\sigma = 20$.

Пример аугментации входных изображений представлен на рис. 7.

5. Результаты экспериментов. Предложенная архитектура сети FusionNet была реализована с использованием среды для машинного обучения TensorFlow на языке программирования Python. Обучение и тестирование сети FusionNet осуществлялось на видеокарте NVIDIA GeForce GTX 1080 с фреймбуфером объемом 8 GB GDDR5X.

Эксперименты на тестовой выборке показали, что в случае хороших погодных условий результат обладает даже большей резкостью, чем результат, выбранный экспертом. Пример результатов комплексирования для изображений, полученных при хороших погодных условиях, представлен на рис. 8.

Эксперименты на тестовой выборке также показали хорошие возможности сети FusionNet в подавлении шумов и устранении артефактов в виде ореолов, возникающих при комплексировании, который использует метод пирамиды Лапласианов. Пример устранения шумов рассмотрен на рис. 9, а пример устранения артефактов в виде ореолов – на рис. 10.

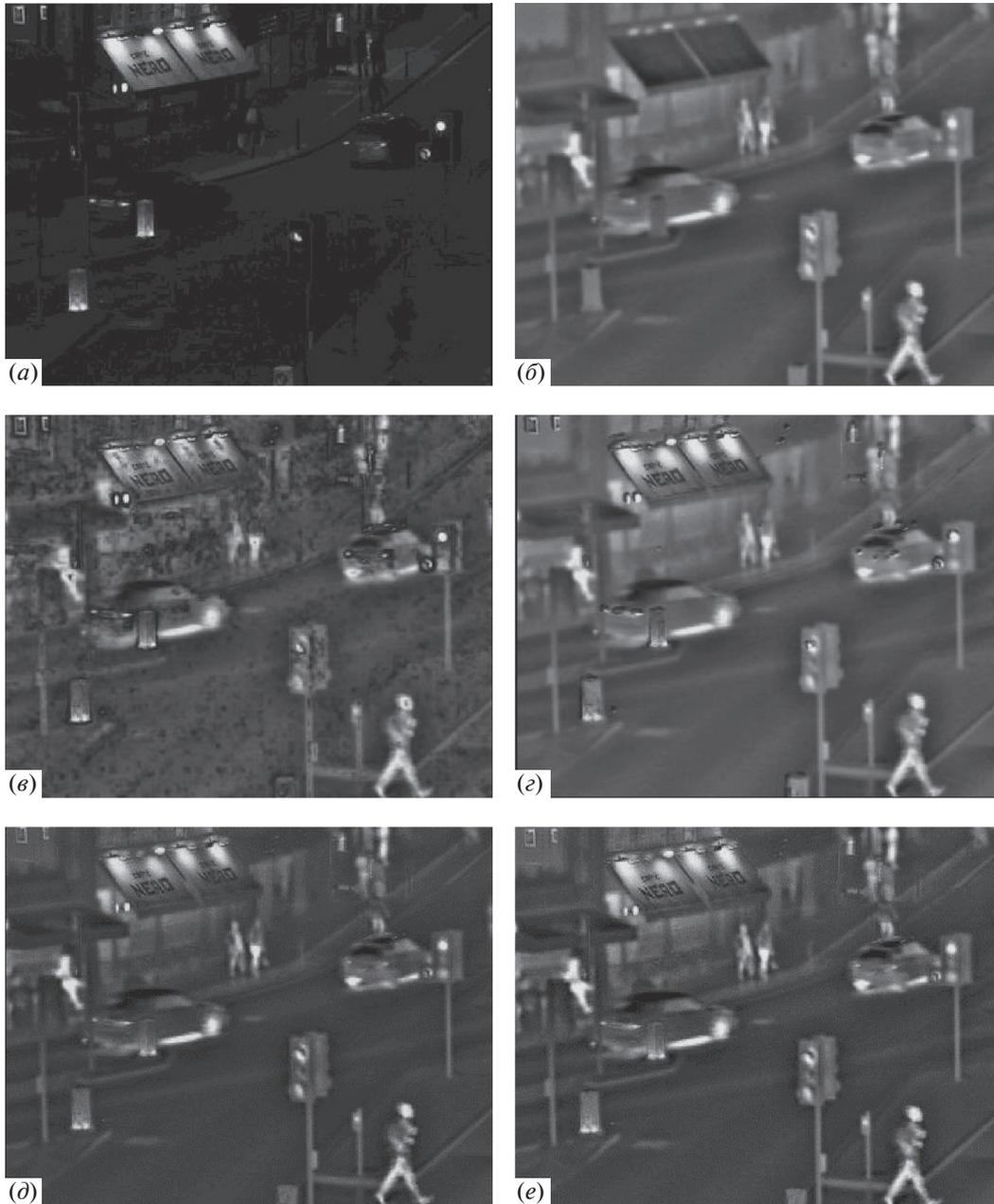


Рис. 13. Примеры результатов разных методов комплексирования: *a* – входное ТВ-изображение, *б* – входное ИК-изображение, *в* – метод CBF, *г* – метод JSR, *д* – метод JSRSD, *е* – метод WLS, *ж* – метод ConvSR, *з* – метод FusionDeepLearn, *и* – наш метод FusionNet

Также сетью FusionNet удачно решается проблема комплексирования в случае сильной инверсии контраста двух разноточных изображений. Пример комплексирования изображений с сильной инверсией контраста на объекте представлен на рис. 11.

В силу того, что во время процесса обучения сети использовались искусственно геометрически рассогласованные изображения, финальная обученная сеть FusionNet осуществляет скрытую внутреннюю юстировку изображений при комплексировании. Пример внутренней юстировки изображений сетью FusionNet описан на рис. 12.

Для сравнения результатов работы сети FusionNet с другими недавними методами комплексирования, использовались изображения и результаты из [11]. Рассматриваются следующие методы комплексирования:

поперечный двусторонний фильтр слияния (CBF) [6];



Рис. 13. Окончание

совместное разреженное представление (JSR) [7];
 JSR метод с обнаружением значимости (JSRSD) [8];
 взвешенный метод наименьших квадратов (WLS) [9];
 сверточное разрежение (ConvSR) [10];
 с использованием глубокого обучения (FusionDeepLearn) [11].

Примеры результатов разных методов комплексирования приведены на рис. 13. Как видно из рис. 13, комплексированное изображение сетью FusionNet не уступает рассматриваемым методам, содержит мало шума, имеет сохранение информативных областей как с ТВ-, так и с ИК-изображений.

Заключение. На основе архитектуры pix2pix предложена оригинальная архитектура сети FusionNet для построения интегрального изображения, содержащего наиболее информативные фрагменты разномасштабных изображений и обладающего большей информативностью по сравнению с каждым из них. Методика построения обучающей и тестовой выборок основана на генерации комплексных изображений различными ранее предложенными методами улучшения и комплексирования разномасштабных данных, сборе экспертных оценок качества результатов комплексирования и обучении генеративной состязательной сети с целью имитации тех вариантов комплексирования, которые получили лучшие экспертные оценки.

Сравнение разработанного метода с другими недавно предложенными методами комплексирования показало, что результаты обученной по такой методике сети FusionNet не уступают ведущим методам, содержат мало шума и обеспечивают сохранение информативных областей со всех входных разномасштабных изображений. Кроме того:

происходит устранение артефактов обработки и комплексирования, вносимых известными методами комплексирования (например, комплексирования на основе пирамиды Лапласианов);

корректно решается проблема информативного комплексирования в случае инверсии контраста двух разнеспектральных изображений;

осуществляется автоматическая внутренняя юстировка изображений в случае небольшого рассогласования входных изображений.

Дальнейшие исследования будут направлены на поиск топологии сети и функции потерь, которая позволила бы получить желаемый результат без заведомо известного эталонного решения.

СПИСОК ЛИТЕРАТУРЫ

1. *Kingma D., Ba J.* Adam: A method for Stochastic Optimization // Intern. Conf. on Learning Representations (ICLR 2015). San Diego, California, USA, 2015. arXiv:1412.6980v9.
2. *Isola P., Jun-Yan Zhu, Zhou T., Efros A.A.* Image-to-Image Translation with Conditional Adversarial Networks // Computer Vision and Pattern Recognition (CVPR 2017). Honolulu, Hawaii, USA, 2017. arXiv:1611.07004.
3. *Ronneberger O., Fischer P., Brox Th.* U-Net: Convolutional Networks for Biomedical Image Segmentation // Computer Vision and Pattern Recognition (CVPR 2015). Boston, Massachusetts, USA, 2015. arXiv:1505.04597.
4. *Jiayi Ma., Yong Ma., Chang Li.* Infrared and Visible Image Fusion Methods and Applications: A survey // Information Fusion. 2019. V. 45. P. 153–178. <https://doi.org/10.1016/j.inffus.2018.02.004>
5. *Инсаров В.В., Тихонова С.В., Михайлов И.И.* Проблемы построения систем технического зрения, использующих комплексирование информационных каналов различных спектральных диапазонов // Новые технологии. Приложение к журналу “Информационные технологии”. 2014. № 3.
6. *Kumar B.K.S.* Image Fusion Based on Pixel Significance Using Cross Bilateral Filter // Signal, Image and Video Processing. 2015. V. 9. I. 5. P. 1193–1204.
7. *Zhang Q., Fu Y., Li H., Jian Z.* Dictionary Learning Method for Joint Sparse Representation-based Image Fusion // Optical Engineering. 2013. V. 52. Iss. 5. <https://doi.org/10.1117/1.OE.52.5.057006>
8. *Liu C.H., Qi Y., Ding W.R.* Infrared and Visible Image Fusion Method Based on Saliency Detection in Sparse Domain // Infrared Physics & Technology. 2017. V. 83. P. 94–102.
9. *Ma J., Zhou Z., Wang B., Zong H.* Infrared and Visible Image Fusion Based on Visual Saliency Map and Weighted Least Square Optimization // Infrared Physics & Technology. 2017. V. 82. P. 8–17.
10. *Liu Y., Chen X., Ward R.K., Wang Z.J.* Image Fusion with Convolutional Sparse Representation // IEEE Signal Processing Letters. 2016. V. 23. Iss. 12. P. 1882–1886.
11. *Li H., Wu X.J., Kittler J.* Infrared and Visible Image Fusion using a Deep Learning Framework // Intern. Conf. on Pattern Recognition (ICPR 2018). Beijing, China, 2018. arXiv:1804.06992.
12. *Petro A.B., Sbert C., Morel J.-M.* Multiscale Retinex // Image Processing On Line. 2014. P. 71–88. <https://doi.org/10.5201/ipol.2014.107>