



Выделение отличий на разносезонных изображениях земной поверхности при помощи генеративных состязательных сетей

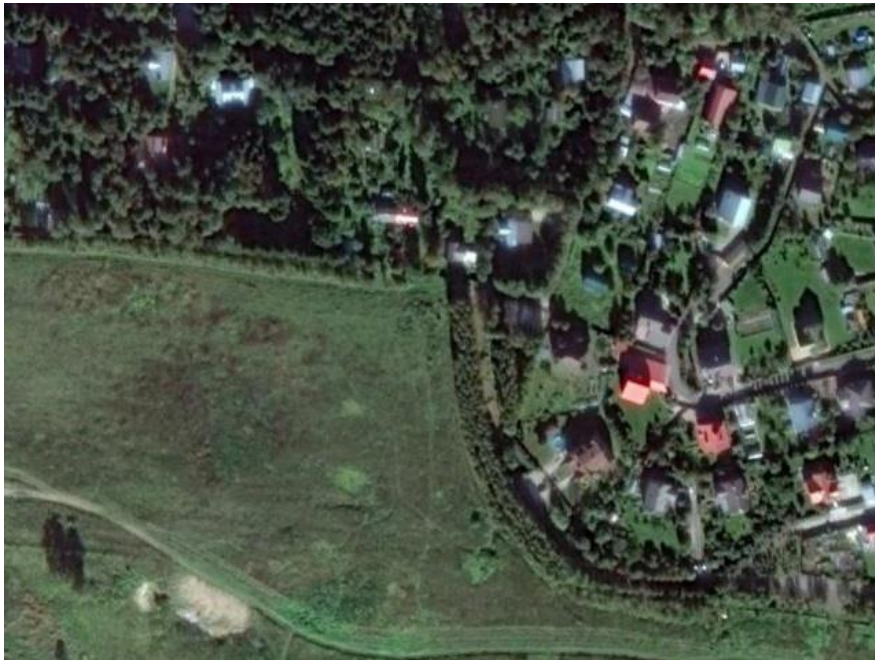
*М.А. Лебедев, Д.В. Комаров,
О.В. Выголов, Ю.В. Визильтер*

MLebedev@gosniias.ru



Особенности задачи выделения отличий

- рассматриваются **разносезонные аэрофотоснимки**;
- **учитываются** изменения, которые соответствуют появлению новых или исчезновению старых объектов;
- **не учитываются** внешние изменения присутствующих объектов, связанных с сезонными изменениями, изменением освещения и другими факторами;
- возможно **рассогласование** изображений **до 5 пкс.**



Особенности задачи выделения отличий

Проблема сопоставления признаков множеств, извлеченных с разных изображений, на сегодняшний день наилучшим образом решается при помощи **генеративно-сопоставительных сетей (GAN)**.

В связи с этим для решения нашей задачи за основу была взята сеть такого типа – условная генеративно-сопоставительная сеть **pix2pix**.



Генеративно-состязательная сеть Pix2Pix

Основными составляющими условной генеративно-состязательной сети являются две конкурирующие нейронные сети: генератор G и дискриминатор D .

Генератор G обучается создавать изображения, которые при помощи конкурирующего дискриминатора D нельзя отличить от «реальных» изображений p_{data} .

$$G = G(z; \theta_g): Z \rightarrow X,$$

где Z – некоторое пространство скрытых (латентных) признаков, на котором задано априорное распределение $p_z(z)$.

Дискриминатор D , в свою очередь, обучается хорошо обнаруживать «поддельные» изображения генератора G . Он выглядит так:

$$D = D(x; \theta_d): X \rightarrow [0,1].$$

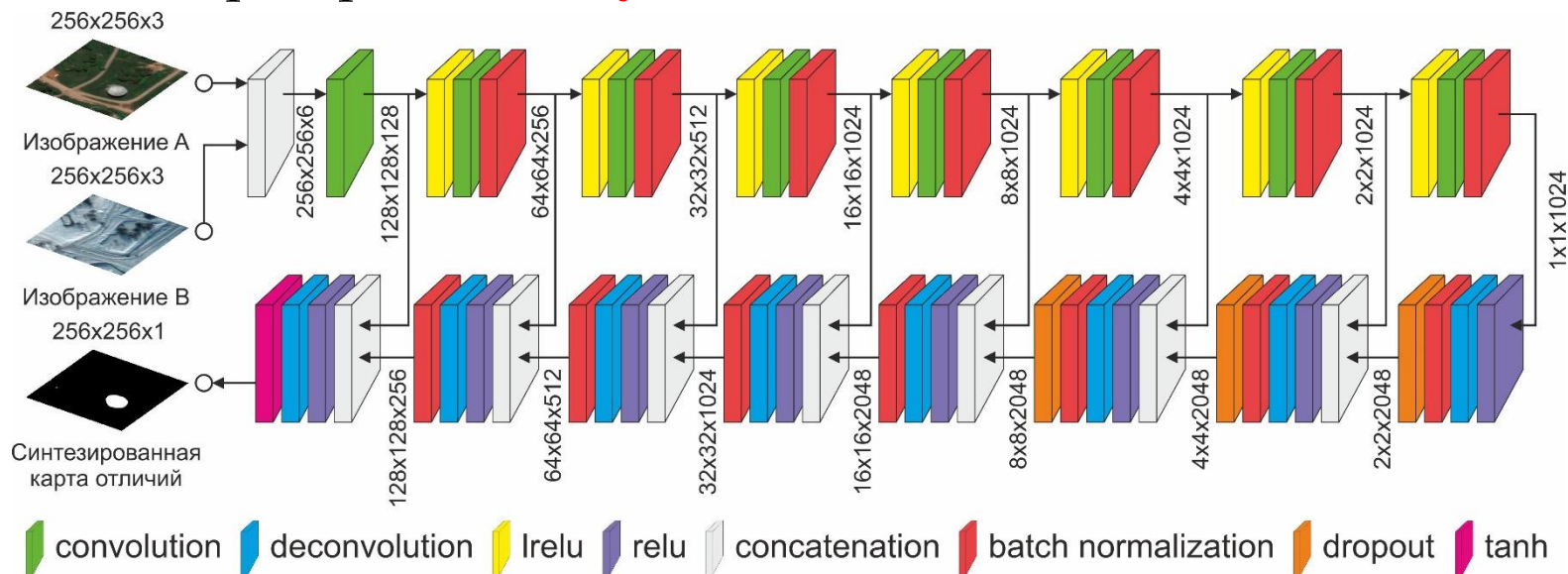
Дискриминатор отображает объекты из пространства данных в отрезок $[0,1]$, который интерпретируется как вероятность того, что пример был действительно «настоящий» из p_{data} . Фактически дискриминатор и генератор играют между собой в игру, которая в теории игр называется минимаксной игрой, решая **следующую оптимизационную задачу**:

$$cGAN = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D), \text{ где}$$

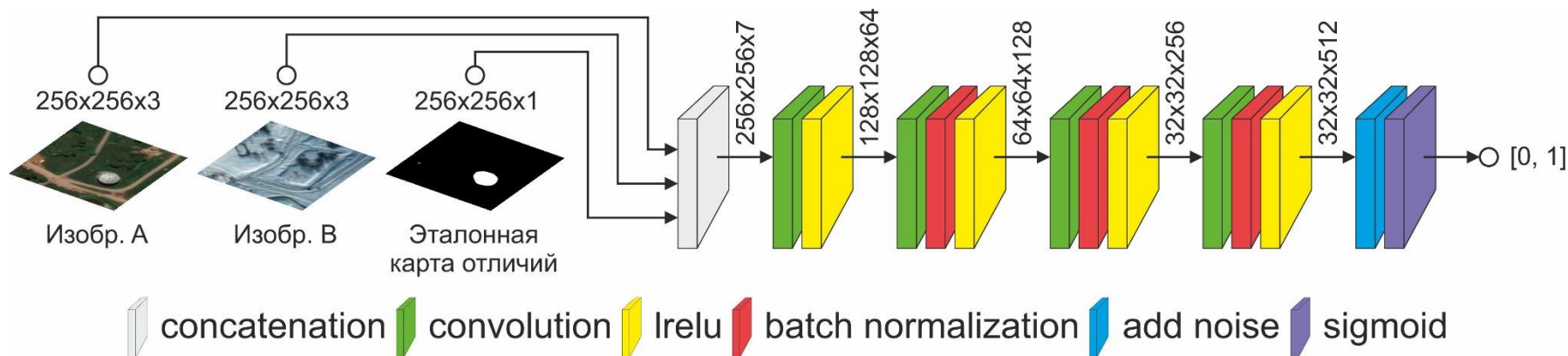
$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y \sim p_{data}(x,y)} [\log D(x, y)] + \mathbb{E}_{x \sim p_{data}(x), z \sim p_z(z)} [\log(1 - D(x, G(x, z)))]$$

Архитектура генератора и дискриминатора

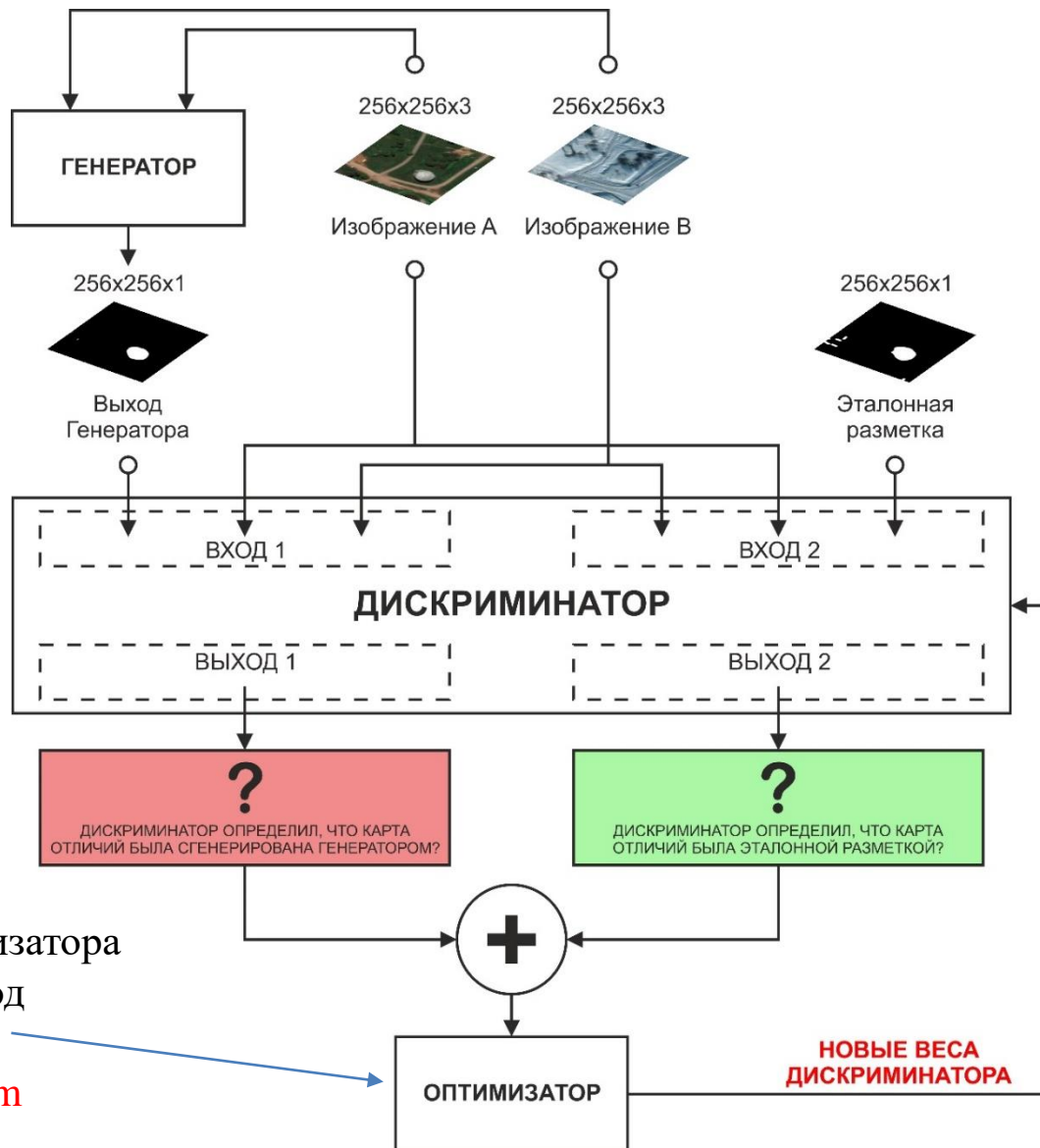
В основе генератора **используется сеть «U-Net»**.



Дискриминатор – это другая отдельная сеть, архитектура которой называется **«PatchGAN»**.

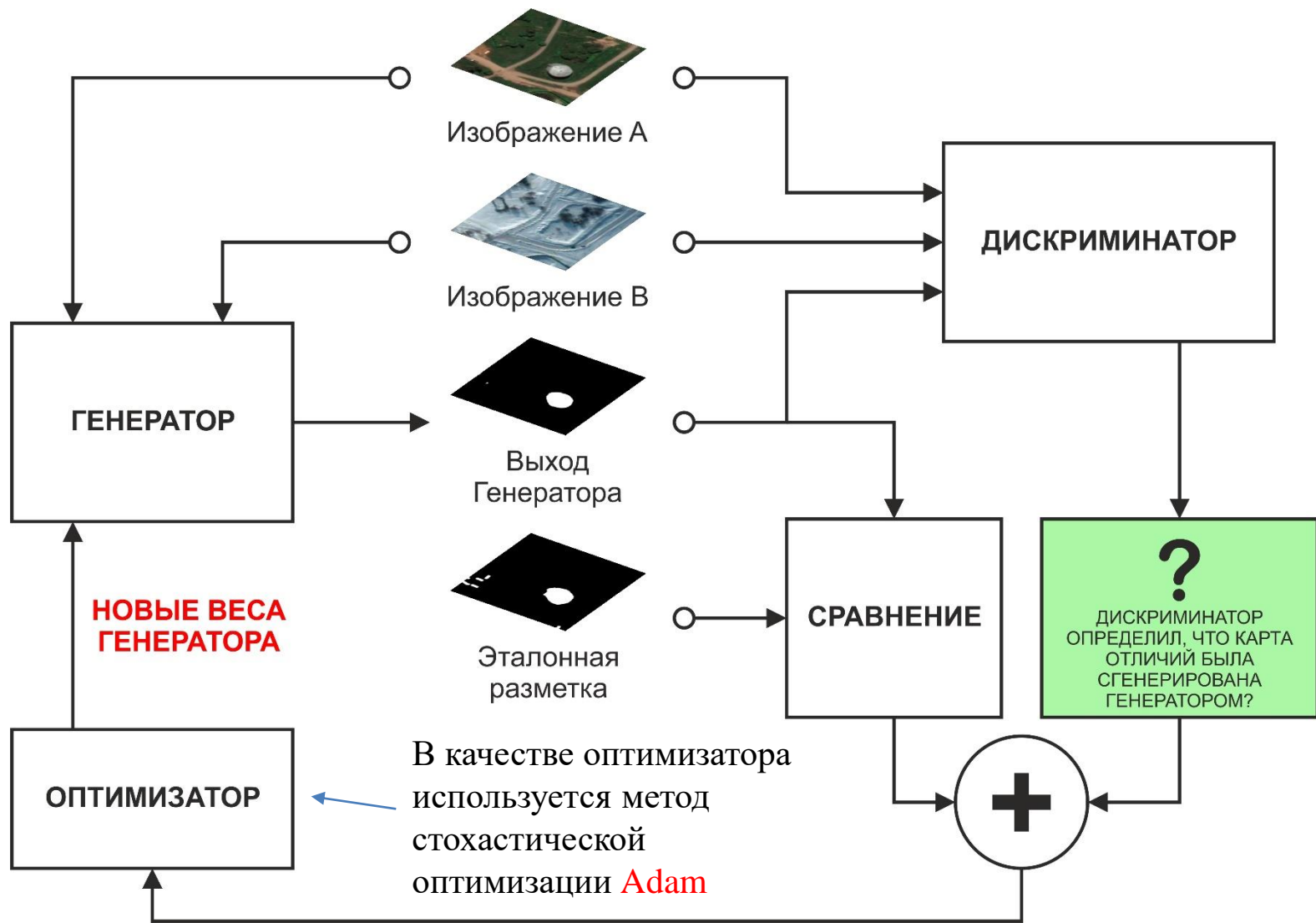


Блок-схема обучения дискриминатора



В качестве оптимизатора используется метод стохастической оптимизации Adam

Блок-схема обучения генератора



Эксперименты

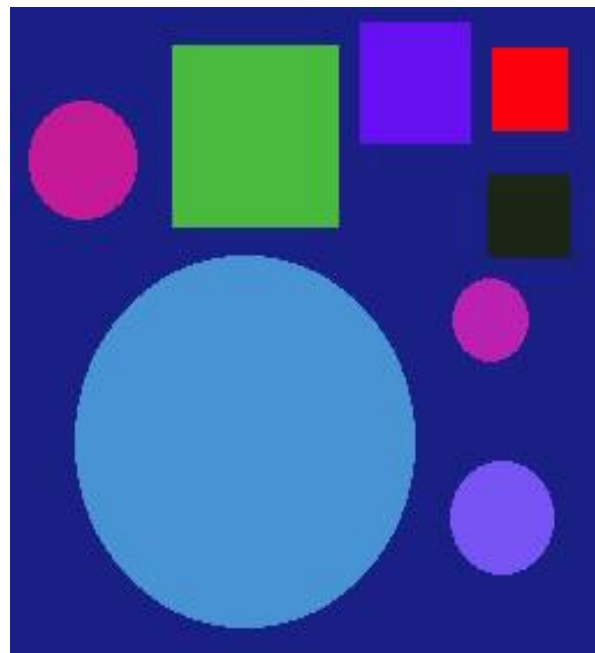
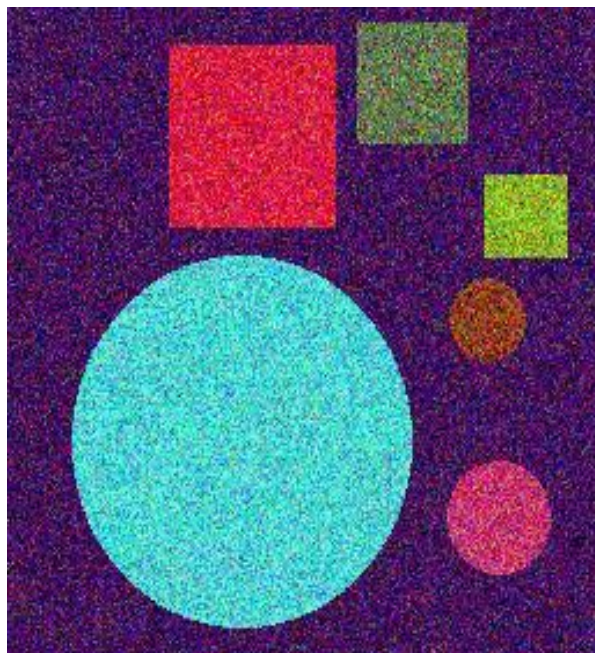
Эксперименты состояли из **трех основных последовательных исследований**:

1. эксперименты на модельных изображениях без относительного сдвига объектов;
2. эксперименты на модельных изображениях с относительным сдвигом объектов;
3. эксперименты на реальных разносезонных изображениях земной поверхности.

Предложенная архитектура сети была реализована при помощи открытой программной библиотеки для машинного обучения **TensorFlow** на языке программирования **Python**. Обучение и тестирование сети осуществлялось на видеокарте **NVIDIA GeForce GTX 1080** с фреймбуфером объемом 8 GB GDDR5X.

Эксперименты на модельных изображениях без сдвига

Создана база из **12 тысяч наборов модельных изображений** размером 256x256 пикселей. База содержала 8 тысяч наборов изображений в качестве обучающей выборки и по 2 тысячи наборов в качестве валидационной и тестовой выборки. Часть изображений подвергалась размытию ($10 < \delta < 25$) и стандартному гауссовскому зашумлению ($10 < \delta < 35$).



Результаты экспериментов на модельных изображениях без сдвига

Для анализа результатов на модельных данных использовались пиксельные показатели точности (*Precision*) и показатели полноты (*Recall*).

Эталонная карта отличий / Сгенерированная сетью карта отличий	Значения пикселя 1	Значения пикселя 0
Значения пикселя 1	True Positive (TP)	False Positive (FP)
Значения пикселя 0	False Negative (FN)	True Negative (TN)

Показатель точности (*Precision*) – доля объектов, названных классификатором положительным и при этом действительно являющимися положительными. Вычисляется по следующей формуле:

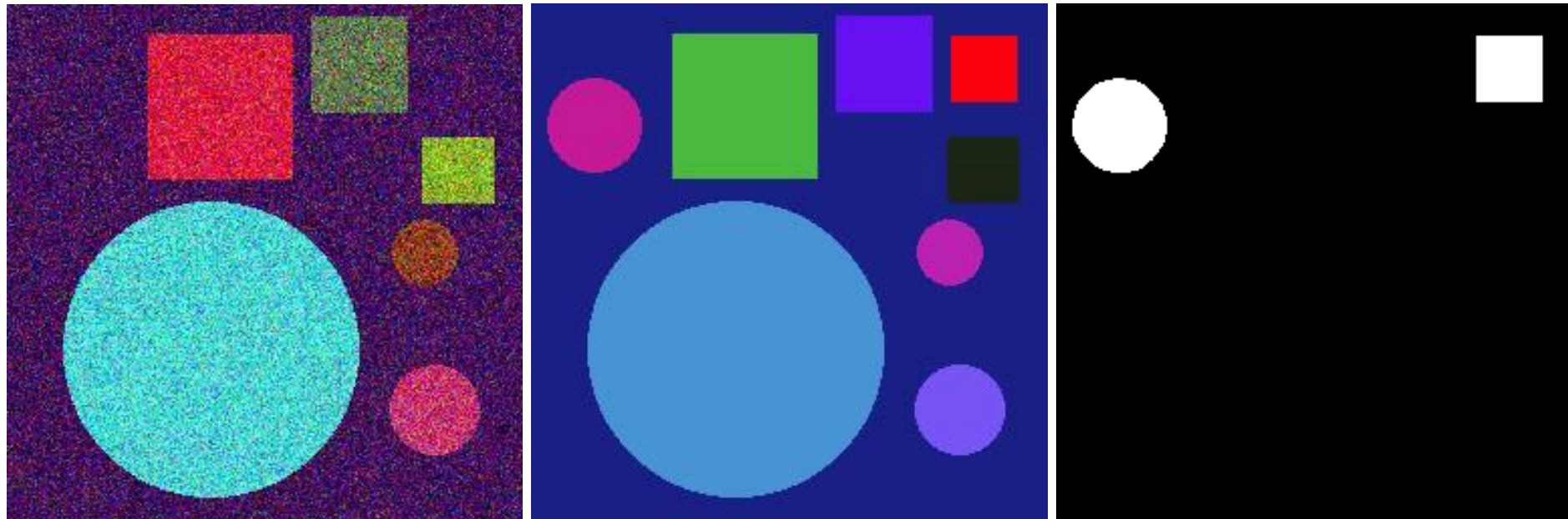
$$Precision = \frac{TP}{TP + FP}$$

Показатель полноты (*Recall*) – доля объектов, названных классификатором положительными, из всех объектов положительного класса. Вычисляется по формуле:

$$Recall = \frac{TP}{TP + FN}$$

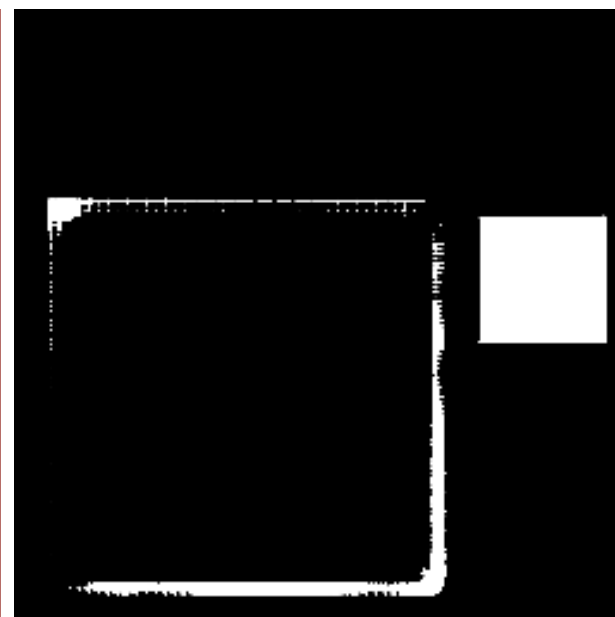
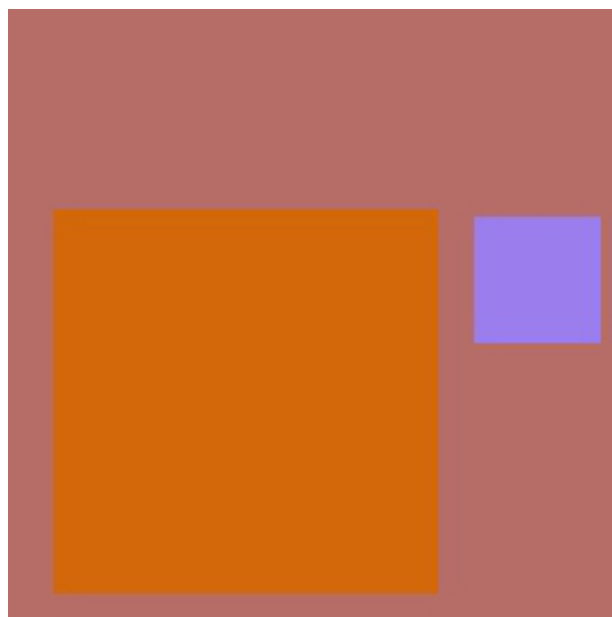
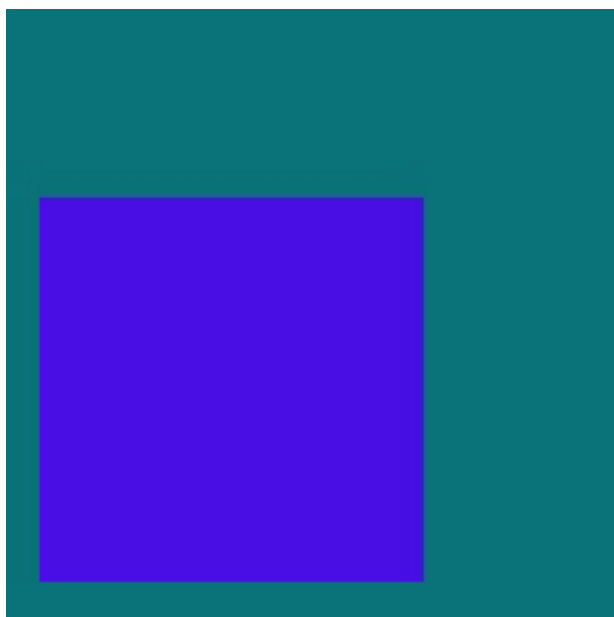
Результаты экспериментов на модельных изображениях без сдвига

В ходе тестирования обученной сети **показатель точности составил 0.95**, а **показатель полноты 0.96**. Наибольший негативный эффект на результат оказывает размытие изображения, а не добавление на него шума.



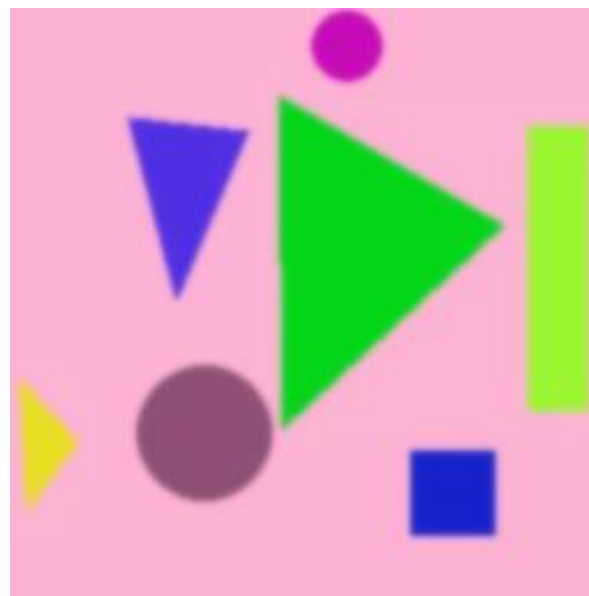
Эксперименты на модельных изображениях со сдвигом

Так как **реальные изображения** земной поверхности, полученные в разное время разными техническими средствами, **могут иметь локальные расхождения**, то на втором этапе исследовалась применимость предложенной архитектуры сети для случая небольших сдвигов объектов.



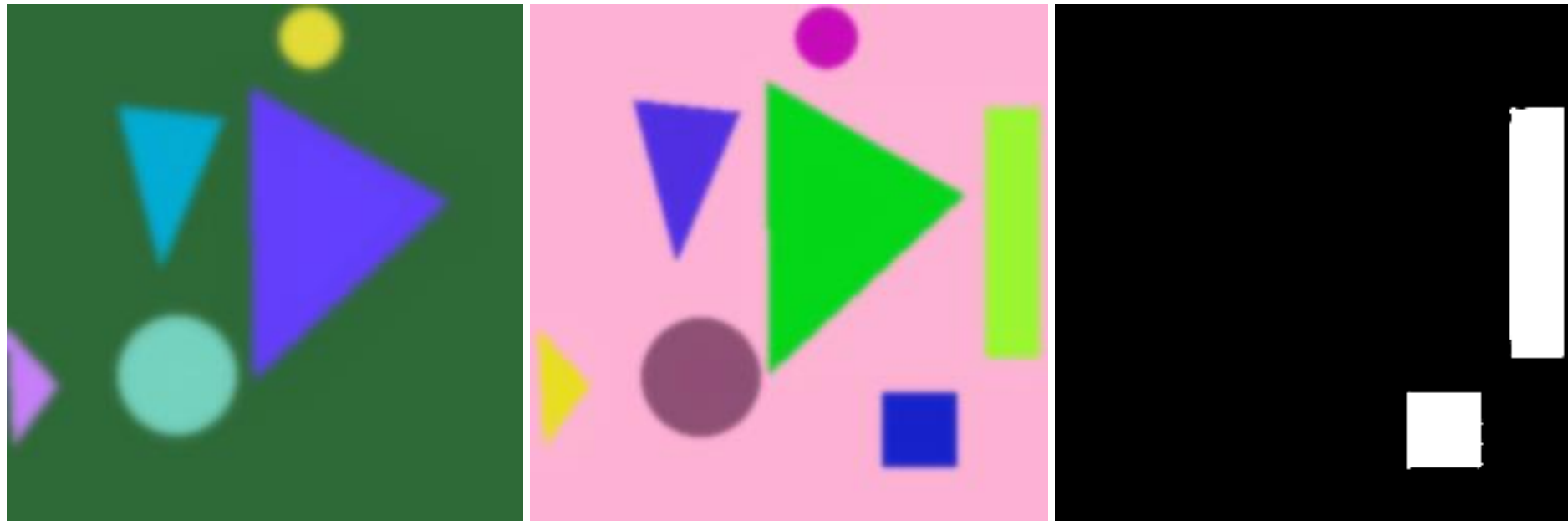
Эксперименты на модельных изображениях со сдвигом

Создана база из **12 тысяч наборов модельных изображений** размером 256x256 пикселей по аналогии с предыдущим методом. Единственным отличием является случайный геометрический сдвиг объектов на изображении В, присутствующих как на изображении А, так и на изображении В. Геометрический сдвиг заключался в случайном индивидуальном смещении объекта от -5 до +5 пикселей по оси абсцисс и ординат без выхода за границы изображения и не пересечения с другими объектами.



Результаты экспериментов на модельных изображениях со сдвигом

В ходе тестирования обученной сети **показатель точности составил (Precision) 0.92**, а **показатель полноты (Recall) 0.93**. Как и в случае с отсутствием относительного сдвига объектов наибольший негативный эффект на результат оказывает размытие изображения, а не добавление на него шума.



Эксперименты на реальных разносезонных изображениях

Для создания базы были использованы разновременные снимки одних и тех же участков местности, выгруженных при помощи программы Google Earth.

- **7 пар разновременных** изображений разрешением **4725x2700** пикселей для ручной разметки
- **4 пары разновременных** изображений разрешением **1900x1000** пикселей с минимальными изменениями для добавления на них модельных изображений.
- Пространственное разрешение выгруженных изображений составляло **от 3 см/px1 до 100 см/px1**, что позволило учитывать как разномасштабные объекты от автомобиля до крупных строительных сооружений, так и сезонные изменения объектов от отдельных деревьев до больших участков лесопосадки.

Примеры выгруженных реальных разносезонных изображениях



Подготовка базы изображений для обучения

База генерировалась путем случайного вырезания фрагмента 256x256 пикселей с поворотом на случайный произвольный угол от 0 до 2π с учетом попадания в вырезаемую область хотя бы части объекта отличий. Итоговая база содержала **16 тысяч наборов изображений** размером 256x256 пикселей: 10 тысяч наборов изображений в качестве обучающей выборки и по 3 тысячи наборов в качестве валидационной и тестовой выборки.



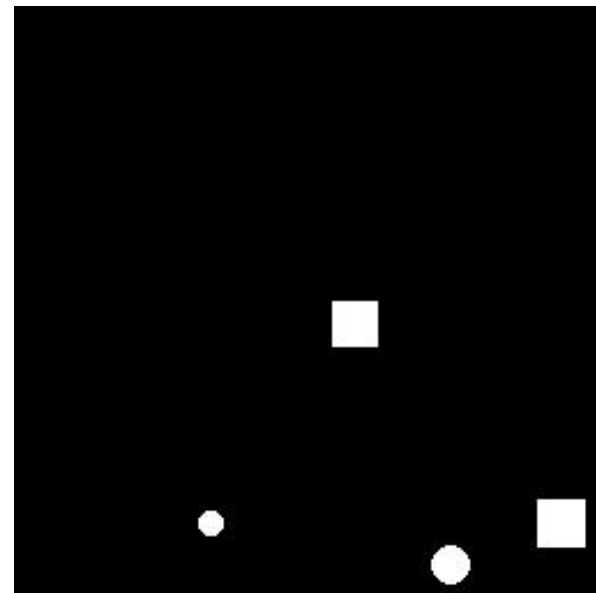
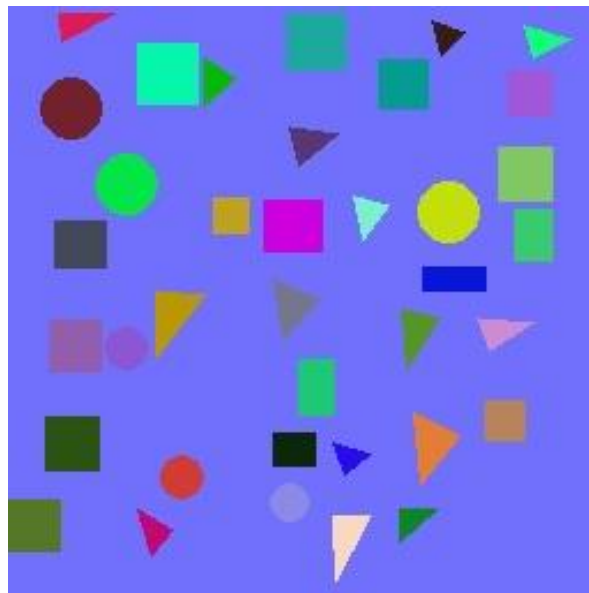
Результаты экспериментов на реальных изображениях

Так как в случае ручной разметки допустимы неточности, то для анализа результатов на реальных изображениях использовалась метрика **Intersection over Union (IoU)**. Для значений порога IoU равным 0.5 по всей выборке средние значения показателей точности (Precision) и полноты (Recall) составили соответственно **0.26** и **0.32**.



Дообучение сети

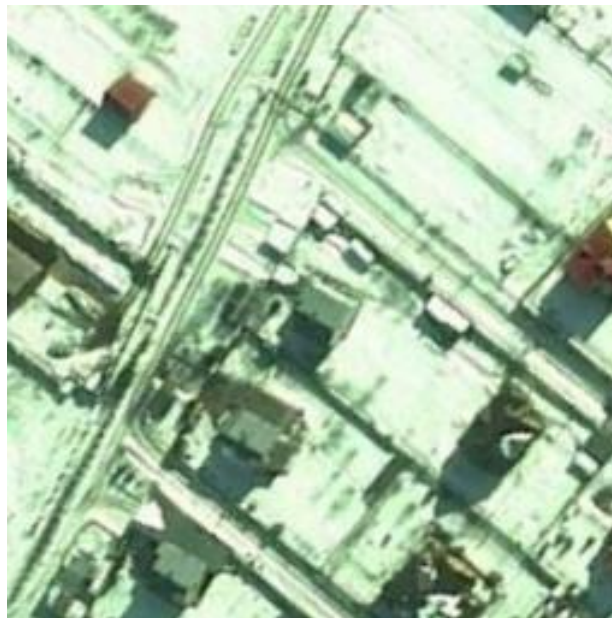
Создана база из **13 тысяч наборов модельных изображений** размером 256x256 пикселей с объектами маленького размера. База содержала 10 тысяч наборов изображений в качестве обучающей выборки и по 3 тысячи наборов в качестве валидационной и тестовой выборки. Часть изображений подвергалась размытию ($10 < \delta < 25$) и стандартному гауссовскому зашумлению ($10 < \delta < 35$).



Результаты экспериментов на реальных изображениях

После дообучения сети результаты средних значений по тестовой выборке показателя точности (Precision) и полноты (Recall) для разных порогов IoU приведены в таблице.

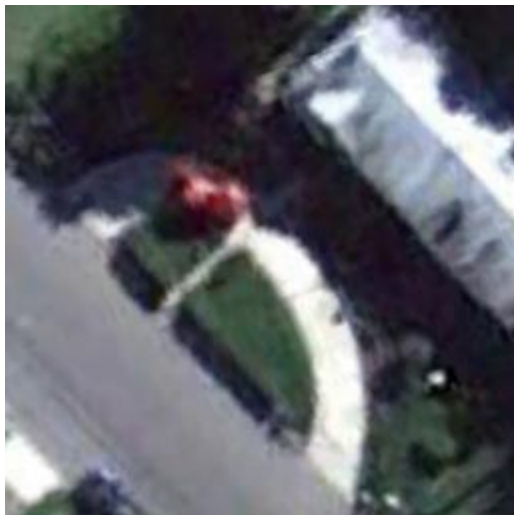
Threshold IoU	Precision	Recall
0.8	0.73	0.63
0.5	0.9	0.84



Результаты экспериментов на реальных изображениях

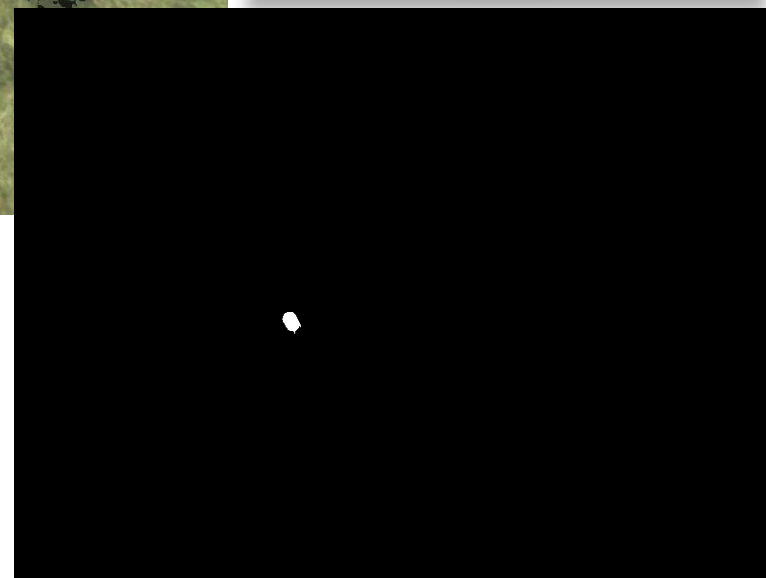
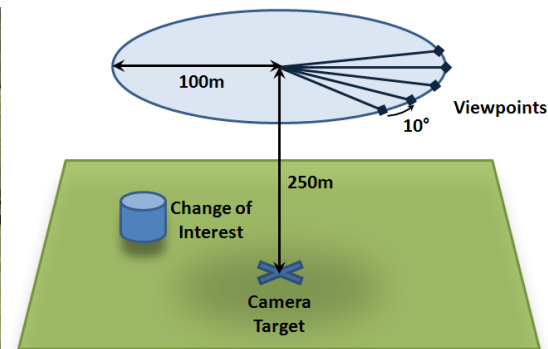


Результаты экспериментов на реальных изображениях



Эксперименты на базе AICDDataset

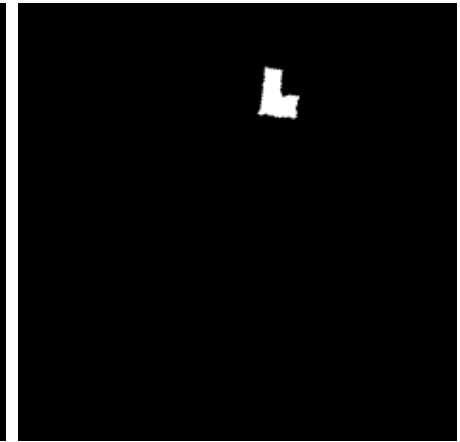
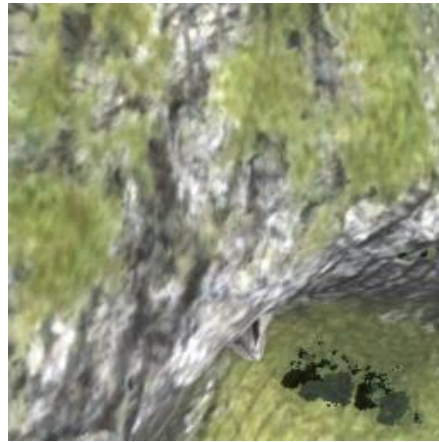
500 пар изображений размером 800x600пкс (100 сцен)



Сгенерировано: 10000 троек на обучение и 1000 троек на тестирование

Результаты экспериментов на базе AICDDataset

В ходе тестирования обученной сети показатель точности составил (Precision) **0.94**, а показатель полноты (Recall) **0.83**.



Заключение

Разработана условная генеративная состязательная сеть для выделения отличий на снимках дистанционного зондирования Земли. Сеть основана на сети pix2pix, оригинальная архитектура которой была модернизирована.

Разработана большая база данных модельных и реальных изображений, содержащая в себе 12 тысяч троек модельных изображений без относительного сдвига объектов, 12 тысяч троек модельных изображений с относительным сдвигом объектов и 16 тысяч троек фрагментов реальных снимков дистанционного зондирования Земли.

В результате проведенных экспериментов предложенный метод выделения отличий подтвердил свою работоспособность на модельных и реальных данных, а также на базе AICDDataset.

Созданная база данных доступна для скачивания по ссылке:

https://drive.google.com/open?id=1GX656JqqOyBi_Ef0w65kDGVto-nHrNs9

Благодарности

Работа выполнена при поддержке РФФ, грант 16-11-00082

Спасибо за внимание!