***Воркшоп №1 «Нейронные сети и их применение в научных исследованиях»***

1. **Название исследовательского проекта:**

***Сегментация и прогнозирование карты облачности на снимках ДЗЗ со спутника ELECTRO L1***

1. **Цели и задачи:**

В данном исследовательском проекте предлагается и рассматривается возможное решение задачи детекции зон облачности, попиксельная привязка к географическим координатам изображений, а также прогнозирование карты облачности по времени со снимков с геостационарного спутника дистанционного зондирования Земли (ДЗЗ) ELECTRO L1, осуществляющего съемку Земли в 10 спектральных каналах. Получение масок облачности и прогноз карты облачности и планируется получить как результат обработки НС.

1. **Описание данных:**

Аппаратура ELECTRO L1 охватывает три канала в спектре VNIR (0.5-0.65, 0.65-0.8, 0.8-0.9 мкм) со спектральным разрешением 1 км для оценки облачного покрова, четыре микроволновых/ИК канала (3.5-4.0, 5.7-7.0, 7.5-8.5, 8.2-9.2 мкм) при пространственном разрешении 4 км для получения изображений в ночное время и измерений водяного пара, три тепловых ИК канала (9.2-10.2, 10.2-11.2, 11.2-12.5 мкм) с пространственным разрешением 4 км для измерения температуры поверхности моря и водяного пара. Таким образом, объект для анализа – изображения в 10 спектральных диапазонах: первые 3 канала – изображения размера 11 136\*11 136 в разрешении 1 км \*1 км, следующие 7 каналов – изображения размера 2 784\*2 784 в разрешении 4 км \* 4км. Периодичность съемки - 30 минут (подробнее про аппаратуру КА и данные с него см. [1]). Также эти же данные есть в более точном представлении в 16-битном формате с исходных файлов L15, предоставляемых напрямую с ПН КА, но требуют самостоятельной дешифровки и обработки для правильного прочтения. Данные со съемочной аппаратуры доступны для получения с сервера научного центра оперативного мониторинга Земли (НЦ ОМЗ). Там доступны для скачивания снимки с ELECTRO L1 за каждые полчаса, начиная с 2012 года по текущий момент времени. Объем данных для анализа очень велик (десятки ТБ). Разбиение на train/val/test должно проводиться на уровне исходных изображений, а не на уровне патчей, чтобы патчи одного изображения целиком находились только в одной из подвыборок.

1. **Метрики и оценки качества, валидация результатов:**

Необходимо осуществлять попиксельную бинарную классификацию облачности по патчам со снимков КА. В качестве оценки достоверности определения маски облачности для каждого пикселя планируется использовать метрику accuracy. В дальнейшей работе возможно использование других метрик, например, F1-score, поскольку некоторые патчи содержат преимущественно облачность, а некоторые – вообще без облаков, также при переходе от бинарной классификации к маскам для нескольких типов облаков их процентное соотношение также будет сильно несбалансированным. При переходе к задаче прогноза карты облачности вероятно будет использована метрика Mean Absolute Error (MAE). Ранее не было никаких исследований и работ с применением НС к данным с ELECTRO L1, однако в других работах по детекции облачности с других КА ДЗЗ основными метриками качества были accuracy и Jaccard Index, причем значение точности бинарной классификации облачности достигалось accuracy = 0.9648, Jaccard Index = 0.785 (см. [2]). Для датасета CloudCast также была использована метрики Frequency Bias и Brier Score (см. [9]).

1. **Модели и алгоритмы:**

Предполагаемая архитектура модели – сверточный автоэнкодер, являющийся разновидностью модели U-Net. Для выделения масок облачности и сверки результата с работой модели НС возможно также использовать не нейросетевые алгоритмы обработки снимков ДЗЗ. При дальнейшем переходе к прогнозированию карты облачности будет осуществлен переход к другой архитектуре - рекуррентным сверточным НС (CNN-LSTM–U-Net).

1. **Результаты:**

На текущем этапе работы сделана предобработка данных с ELECTRO L1 под формат, схожий с изображениями с LANDSAT-8: изображения Земли 11 136\*11 136 пикселей разделены на патчи размера 384\*384 пикселей, вычленены 4 спектральных канала в видимом и ближнем инфракрасном диапазонах из 3 каналов с ELECTRO L1 в соответствии с каналами на съемочной аппаратуре LANDSAT-8, создана базовая модель (baseline) U-Net для разметки облачности для каждого патча с объединением в единую маску для всего снимка Земли 11 136 на 11 136 пикселей, тем самым возможно получение собственного размеченного датасета для данных с КА ELECTRO L1, структура которого схожа с датасетом 38-Cloud с KA LANDSAT-8, также найдены и исследованы 4 датасета с других КА ДЗЗ: 38-Cloud [7], 95-Cloud [6] (расширение для датасета 38-Cloud), SPARCS [5] и CloudCast [8], имеющие схожую тематику данных, они также будут использоваться для дальнейшей работы.

1. **Список литературы и описание датасетов и ресурсов:**
2. Официальный сайт научного центра оперативного мониторинга Земли (НЦ ОМЗ): <https://new.ntsomz.ru/elektro/>.
3. Sorour Mohajerani, Parvaneh Saeedi. CLOUD-NET: AN END-TO-END CLOUD DETECTION ALGORITHM FOR LANDSAT 8. School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada. <https://arxiv.org/pdf/1901.10077.pdf>.
4. Документация Pytorch: <https://pytorch.org/tutorials/>
5. S. Mohajerani and P. Saeedi, "Shadow Detection in Single RGB Images Using a Context Preserver Convolutional Neural Network Trained by Multiple Adversarial Examples" in IEEE Transactions on Image Processing, vol. 28, no. 8, pp. 4117-4129, Aug. 2019, doi: 10.1109/TIP.2019.2904267.
6. Описание датасета SPARCS: <https://www.usgs.gov/landsat-missions/spatial-procedures-automated-removal-cloud-and-shadow-sparcs-validation-data>
7. Описание датасета 95-Cloud: <https://github.com/SorourMo/95-Cloud-An-Extension-to-38-Cloud-Dataset>
8. Описание датасета 38-Cloud: <https://github.com/SorourMo/38-Cloud-A-Cloud-Segmentation-Dataset>
9. Описание датасета CloudCast: <https://www.kaggle.com/datasets/christianlillelund/the-cloudcast-dataset-small>
10. Nielsen, A. H., Iosifidis, A., & Karstoft, H. (2021). CloudCast: A Satellite-Based Dataset and Baseline for Forecasting Clouds. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 3485–3494. doi:10.1109/jstars.2021.306293
11. Hughes, M., & Hayes, D. (2014). Automated Detection of Cloud and Cloud Shadow in Single-Date Landsat Imagery Using Neural Networks and Spatial Post-Processing. Remote Sensing, 6(6), 4907–4926. doi:10.3390/rs6064907