

Camera-based Water Stage and Discharge Prediction with ML

Rodrigo Morales Aguayo A01632834

Jessica Nicole Copado Leal A01637876

Andrés Olvera Rodríguez A01638129

Carlos Estrada Ceballos A01638214



**Tecnológico
de Monterrey**



Introduction

Find a way to predict the state of water (stage and discharge) at any time in the future using ML, sensor data and photographs of water bodies.

Steps taken

- Understand research paper
- Analyze dataset
- Select features
- Clean data
- Create models
- Separate data by seasons
- Use images to create different models



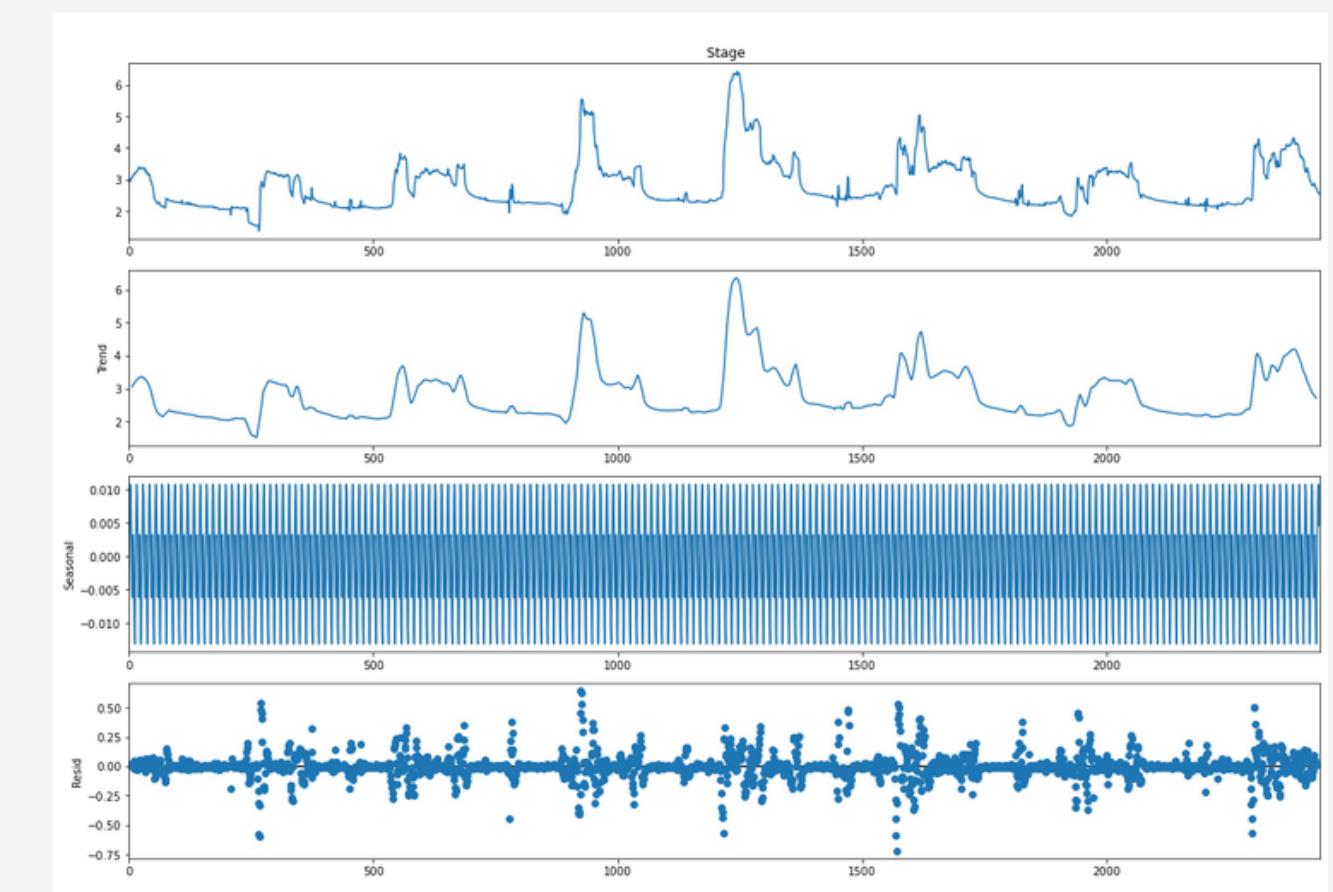
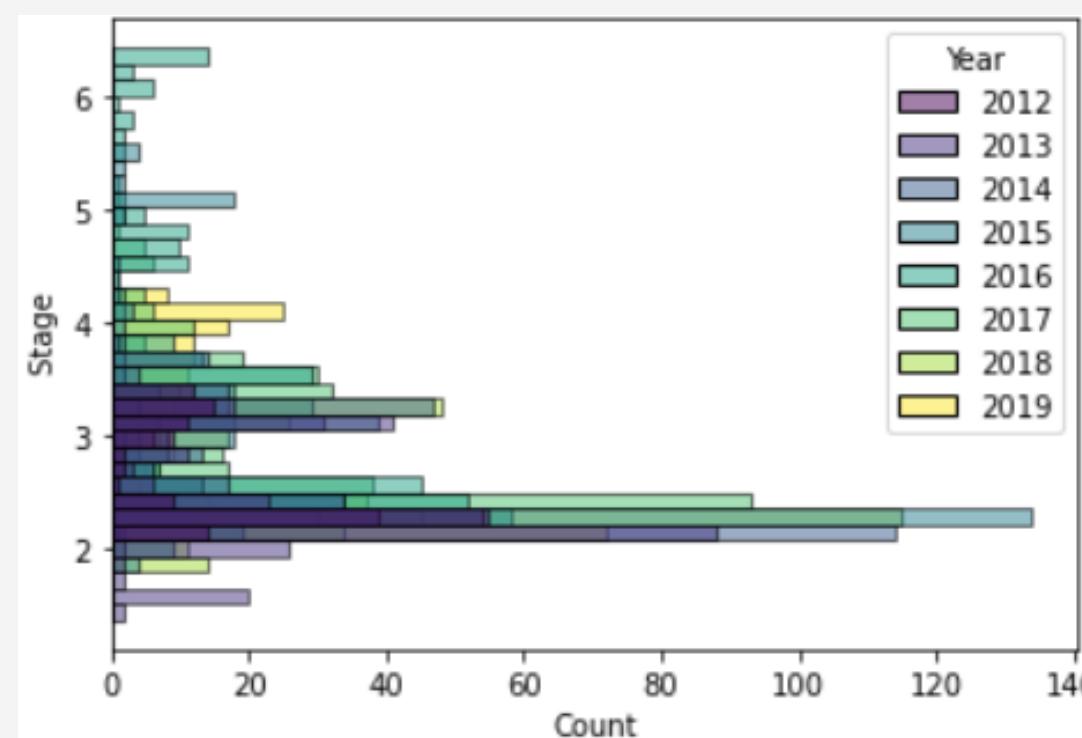
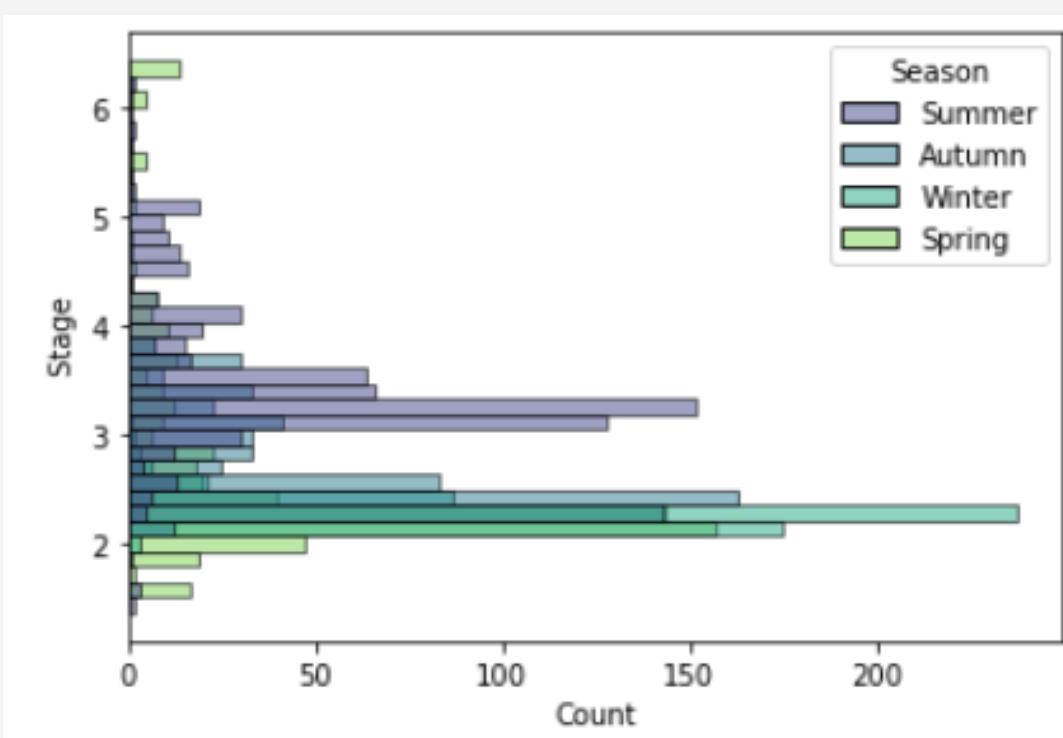
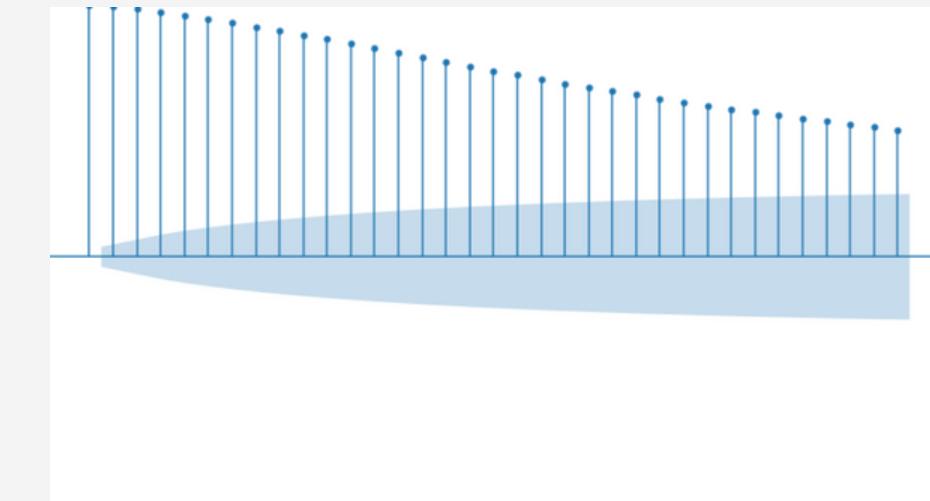
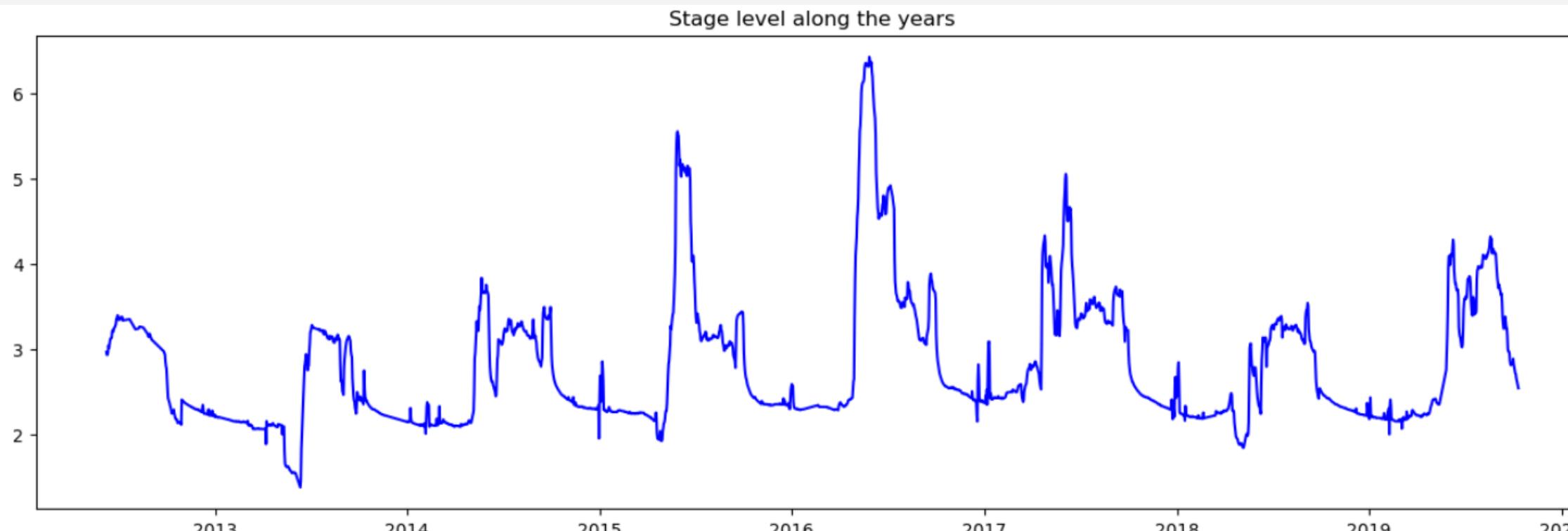
Time series analysis

For our time series analysis we take into account only the *CaptureTime* and the *Stage*. Because there were some records with the same date, we grouped the dates and obtained the mean from each date.

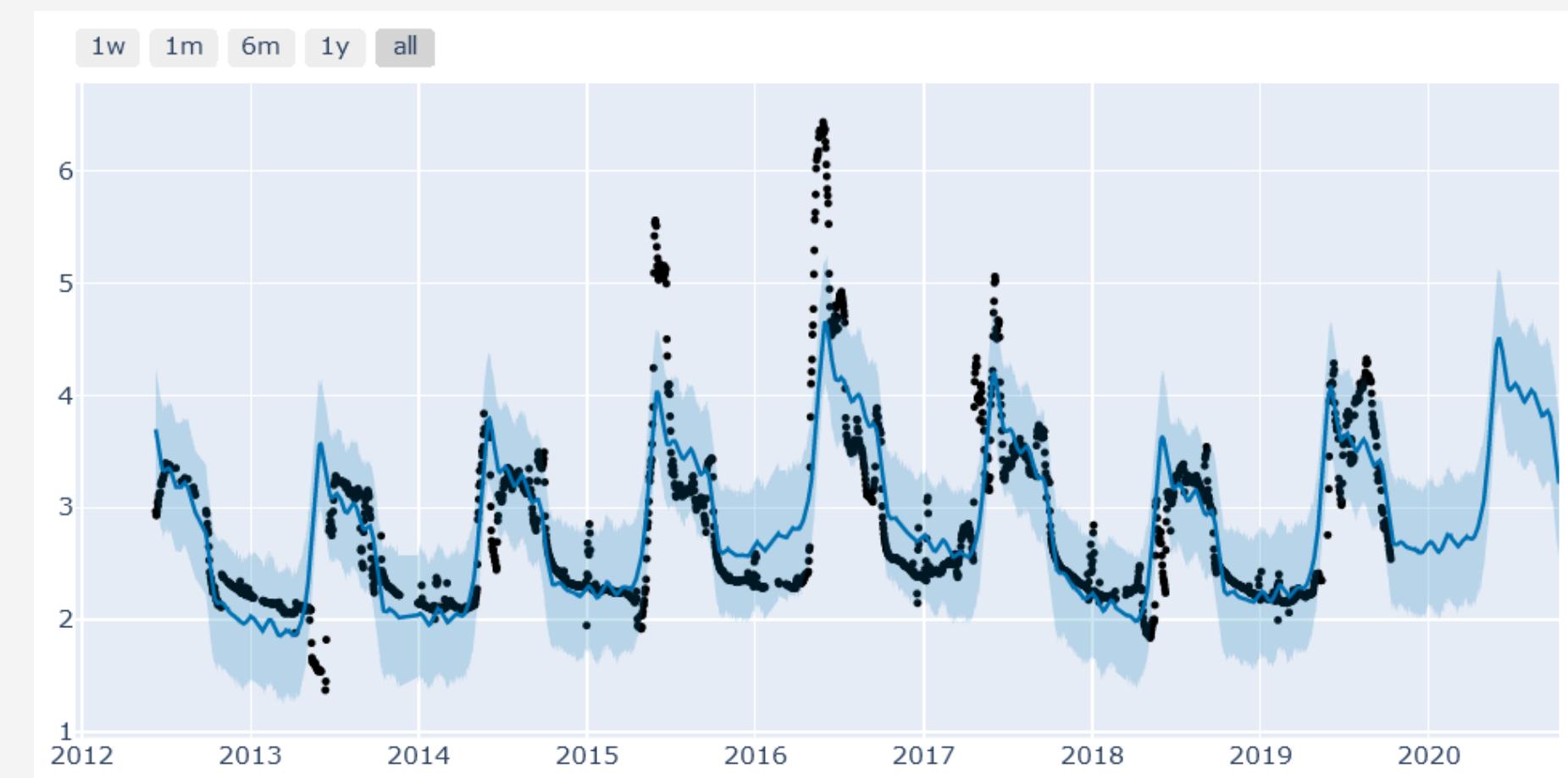
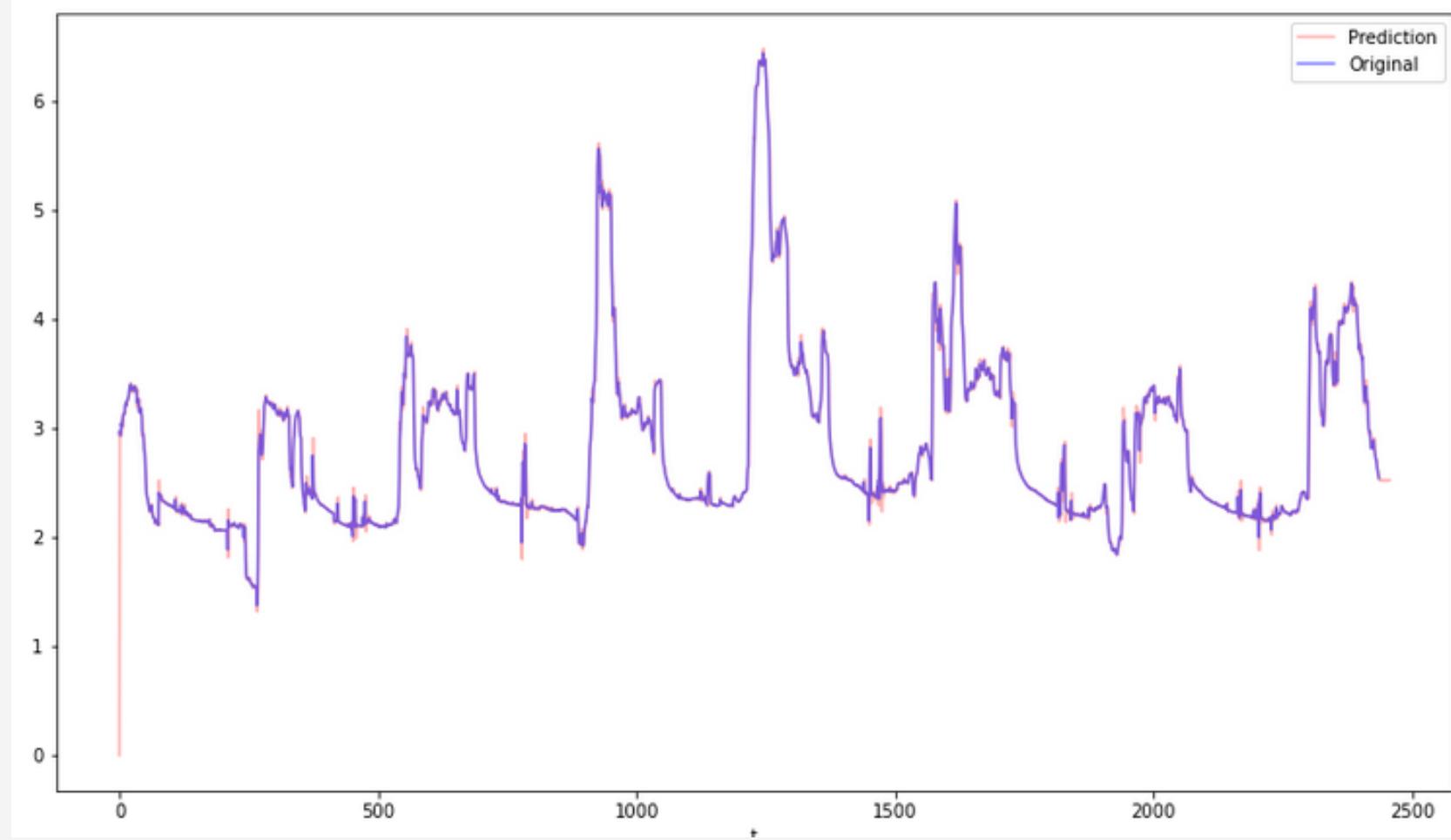
| | CaptureTime | Stage |
|---|-------------|-------|
| 0 | 2012-06-09 | 2.99 |
| 1 | 2012-06-09 | 2.99 |
| 2 | 2012-06-09 | 2.96 |
| 3 | 2012-06-09 | 2.94 |
| 4 | 2012-06-09 | 2.94 |

| | |
|------------|----------|
| 2012-06-09 | 2.965455 |
| 2012-06-10 | 2.931176 |
| 2012-06-11 | 2.931875 |
| 2012-06-12 | 2.965625 |
| 2012-06-13 | 3.042500 |

EDA



ARIMA and Prophet models



Predicting water stage level with regression models

Linear Regression model

With the dataframe provided, we built a linear regression model in order to predict the water stage. We are still trying to determine what is causing the models overfitting. The next thing to do is to use some time series analysis in the variables if we can detect which variables have tendency.

First approach

```
OLS Regression Results
=====
Dep. Variable: Stage    R-squared:      0.663
Model:           OLS     Adj. R-squared:  0.663
Method:          Least Squares   F-statistic:   1286.
Date:  Mon, 28 Nov 2022 Prob (F-statistic): 0.00
Time:       14:14:27   Log-Likelihood: -19302.
No. Observations: 29441   AIC:            3.870e+04
Df Residuals:    29395   BIC:            3.908e+04
Df Model:        45
Covariance Type: nonrobust
=====
```

Second approach

```
OLS Regression Results
=====
Dep. Variable: Stage    R-squared:      0.672
Model:           OLS     Adj. R-squared:  0.671
Method:          Least Squares   F-statistic:   1368.
Date:  Thu, 01 Dec 2022 Prob (F-statistic): 0.00
Time:       14:05:49   Log-Likelihood: -20730.
No. Observations: 28811   AIC:            4.155e+04
Df Residuals:    28767   BIC:            4.191e+04
Df Model:        43
Covariance Type: nonrobust
=====
```

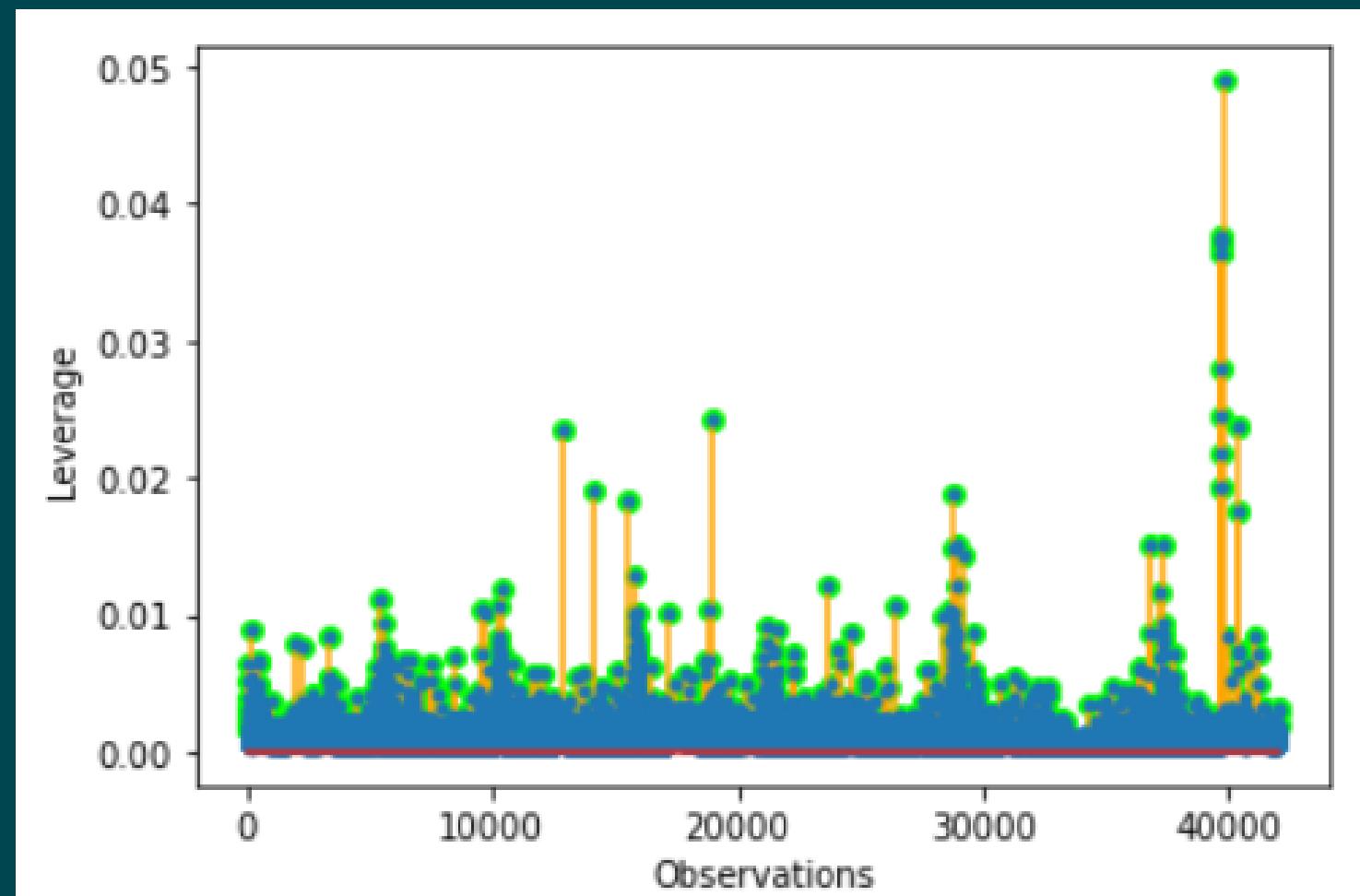
Outlier removal

Once the variables with p-value higher than 0.05 were deleted, we proceeded to the outlier removal.

We used three different ways to detect and eliminate outliers:

- Studentized residuals
- Leveraging
- Cook distance

All three at the same time.



Final model first approach

Decision Tree Regressor scores

r2 score: 0.8252936823154983

Mean squared error (MSE): 0.07423337253189911

Root mean squared error (RMSE): 0.07423337253189911

Mean absolute error (MAE): 0.1578338527243486

Random Forest Regressor scores

r2 score: 0.9118831584291581

Mean squared error (MSE): 0.03744117793424774

Root mean squared error (RMSE): 0.03744117793424774

Mean absolute error (MAE): 0.1089957098760697

Support Vector Regressor scores

r2 score: 0.3048800757597723

Mean squared error (MSE): 0.2953590744420336

Root mean squared error (RMSE): 0.2953590744420336

Mean absolute error (MAE): 0.3969163167570166

MLP Regressor scores

r2 score: -1091.1091901902423

Mean squared error (MSE): 464.04130906878333

Root mean squared error (RMSE): 464.04130906878333

Mean absolute error (MAE): 19.270464385486186

Final model second approach

Decision Tree Regressor scores

r2 score: 0.3438572335928919

Mean squared error (MSE): 0.25310834649694347

Root mean squared error (RMSE): 0.25310834649694347

Mean absolute error (MAE): 0.34076107139304734

Random Forest Regressor scores

r2 score: 0.5519599760739231

Mean squared error (MSE): 0.1728323094093502

Root mean squared error (RMSE): 0.1728323094093502

Mean absolute error (MAE): 0.29036569295015197

Support Vector Regressor scores

r2 score: 0.05946129616394302

Mean squared error (MSE): 0.36281463171174855

Root mean squared error (RMSE): 0.36281463171174855

Mean absolute error (MAE): 0.45964094953648577

MLP Regressor scores

r2 score: -1637.8183113941418

Mean squared error (MSE): 632.177346520528

Root mean squared error (RMSE): 632.177346520528

Mean absolute error (MAE): 16.035323478837054

OLS score

r2 score: 0.801

MSE: 0.084

RMSE: 0.290

MAE: 0.217

OLS score

r2 score: 0.801

MSE: 0.206

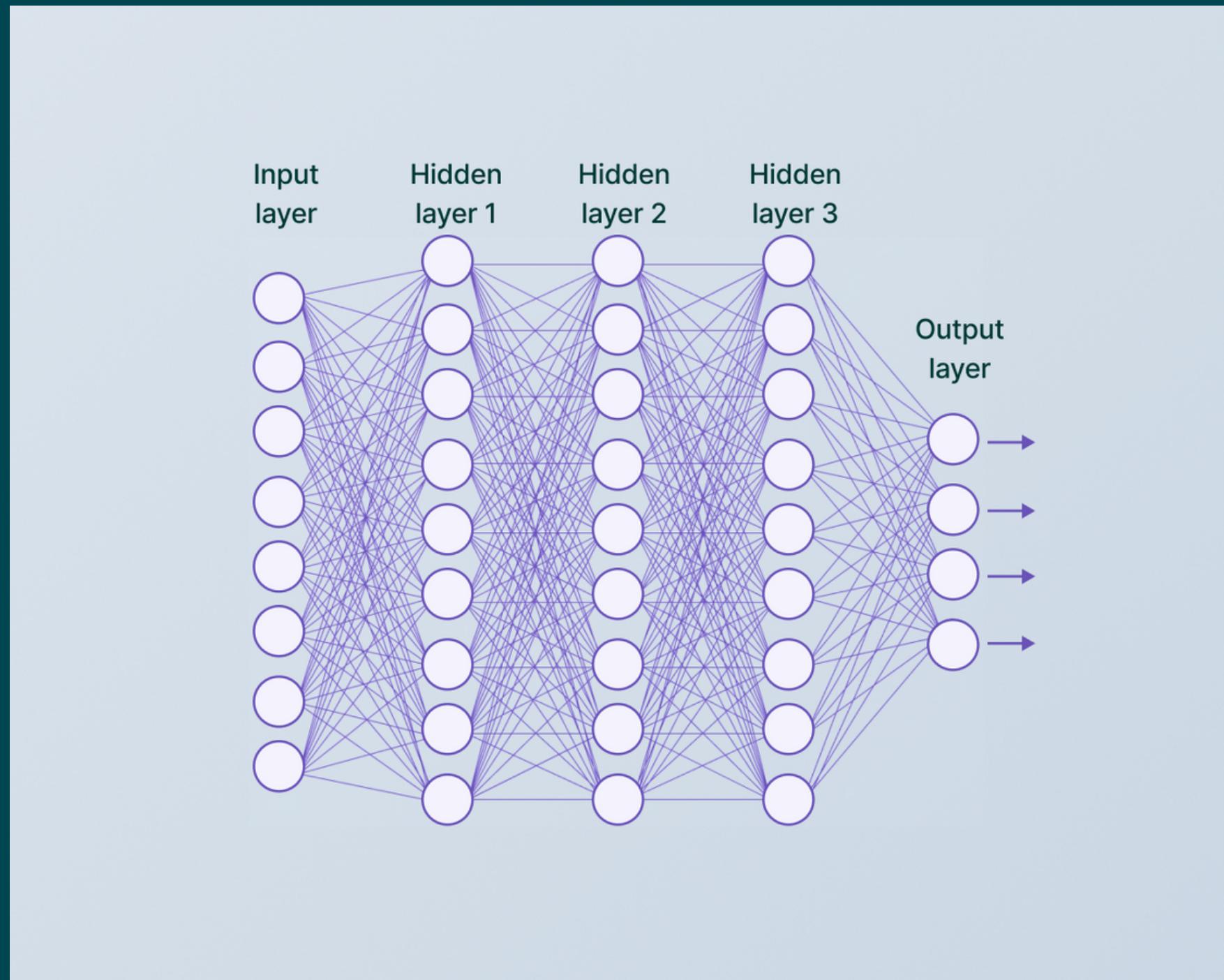
RMSE: 0.454

MAE: 0.340

Images



CNN





Original images

| Train | Test | Validation |
|--------|--------|------------|
| 25,235 | 12,618 | 4,206 |



| MSE | RMSE | MAE |
|-------|------|------|
| 0.123 | 0.35 | 0.22 |

Cropping (river edges)

| Train | Test | Validation |
|--------|--------|------------|
| 25,235 | 12,618 | 4,206 |



| MSE | RMSE | MAE |
|-------|------|-------|
| 0.207 | 0.45 | 0.313 |

Cropping (white water)

| Train | Test | Validation |
|--------|--------|------------|
| 25,235 | 12,618 | 4,206 |

| MSE | RMSE | MAE |
|-------|-------|-------|
| 0.158 | 0.398 | 0.253 |



Split data by year

Spring

Summer

Fall

Winter



Data by year

| Train | Test | Validation |
|--------|--------|------------|
| 21,421 | 15,242 | 5,396 |

Train: 2012, 2013, 2014, 2015, 2016

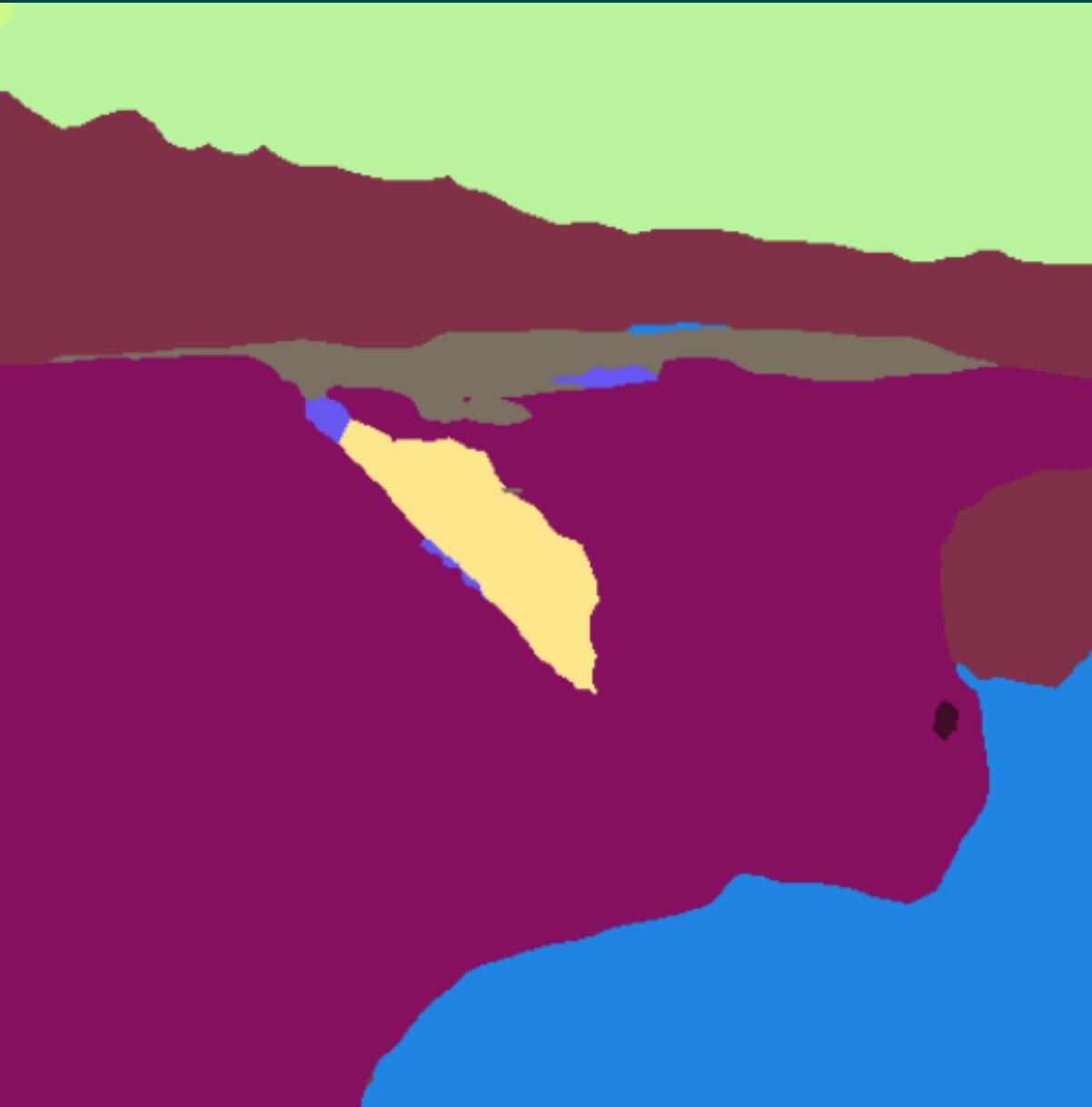
Test: 2017, 2018

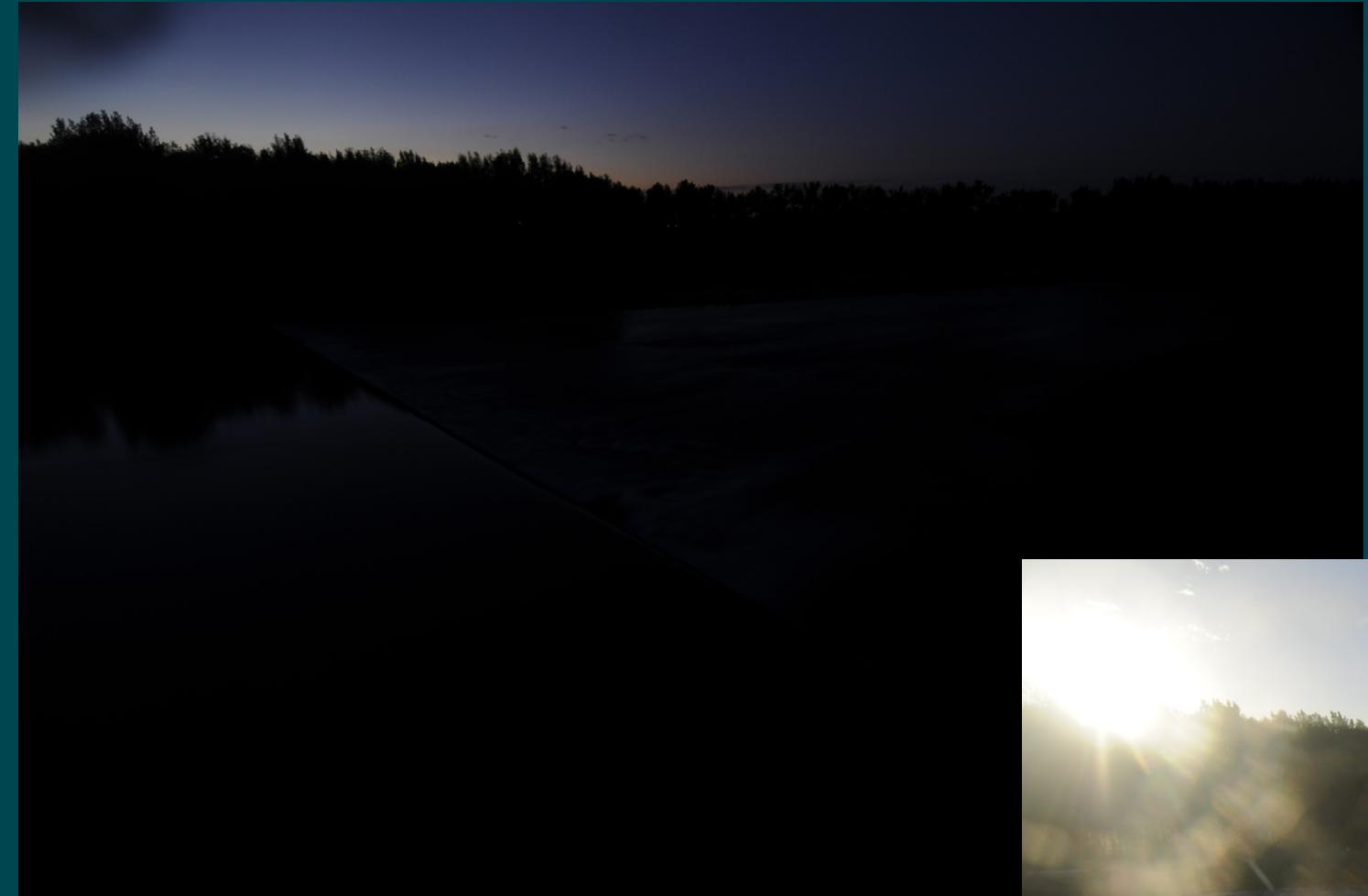
Validation: 2019

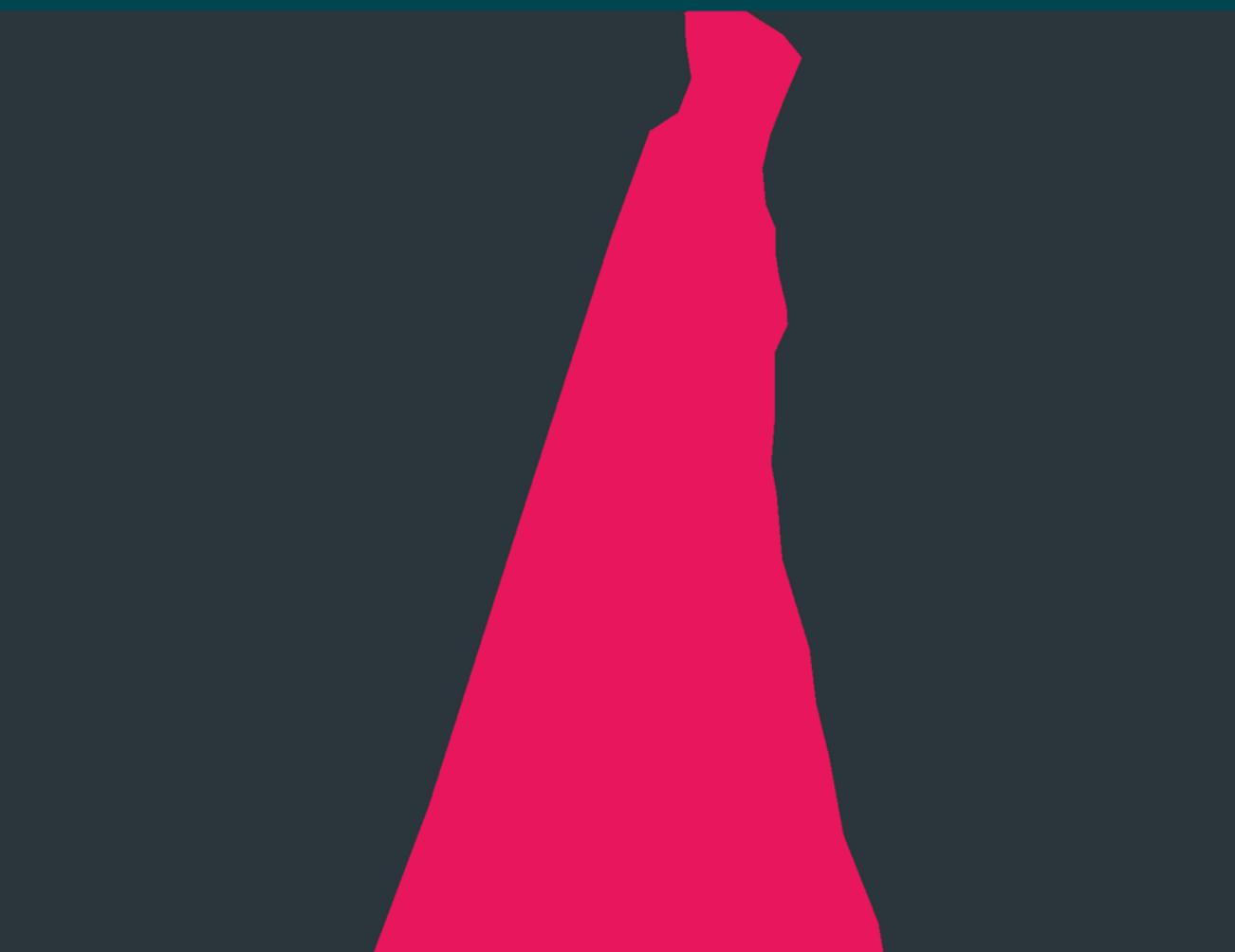


| MSE | RMSE | MAE |
|------|------|------|
| 0.27 | 0.52 | 0.34 |

Segmentation







A two-stage image process for water level recognition via dual-attention CornerNet and CTransformer

Run Qiu, Zhaoxi Cai , Zhuoqing Chang, Shubo Liu & Guoqing Tu

The Visual Computer (2022) | Cite this article

203 Accesses | 1 Altmetric | Metrics

Abstract

Image processing-based water level detectors have promising practical application value in intelligent agriculture and early water logging alerts. However, water level recognition based on image processing faces illumination, shooting angle, and sediment contamination challenges. In addition, due to the influence of water surface reflection, it is not easy to extract the water level ruler (WLR) on the water surface accurately. This paper proposes a novel dual-attention CornerNet for WLR image extraction and CTransformer for WLR sequence recognition. First, a dual-attention mechanism to obtain the global information is introduced to better predict semantic segmentation feature maps and corner information. Then, asymmetric convolution Resnet-50 is used to extract multi-local information to effectively recognize inconsistent character sizes caused by different shooting angles of WLRs. Recently, the design of vision backbone using self-attention becomes an exciting topic. In this work, an improved CTransformer is designed to retain sufficient global context information and extract more differentiated features for sequence recognition via multi-head self-attention. Evaluation using our in-house dataset shows that the proposed framework achieves an F-score of 91.37 in the detection stage and the accuracy of human estimation error within 0.3 cm in the recognition stage is 95.37%, respectively. The proposed method is also evaluated on several benchmarks. Experiment results demonstrate that the method in this paper is superior to the existing methods.

Article

Research on Water-Level Recognition Method Based on Image Processing and Convolutional Neural Networks

Gang Dou ^{1,2}, Rensheng Chen ^{1,3,*}, Chunyan Han ^{1,2}, Zhangwen Liu ¹ and Junfeng Liu ¹

Abstract: Water level dynamics in catchment-scale rivers is an important factor for surface water studies. Manual measurement is highly accurate but inefficient. Using automatic water level sensors has disadvantages such as high cost and difficult maintenance. In this study, a water level recognition method based on digital image processing technology and CNN is proposed. For achieving batch segmentation of source images, the coordinates of the water ruler region in the source image and characters' region and the scale lines' region on the ruler are obtained by using image processing algorithms such as grayscale processing, edge detection, and the tilt correction method based on Hough-transform and morphological operations. The CNN is then used to identify the value of digital characters. Finally, the water level value is calculated according to the mathematical relationship between the number of scale lines detected by pixel traversal in the binarized image and the value of digital characters. This method is used to identify the water levels of the water ruler images collected in the Hulu watershed of the Qilian Mountains in Northwest China. The results show that the accuracy compared with the actual measured water level reached 94.6% and improved nearly 24% compared to the template matching algorithm. With high accuracy, low cost, and easy deployment and maintenance, this method can be applied to water level monitoring in mountainous rivers, providing an effective tool for watershed hydrology research and water resources management.

Qiu, R., Cai, Z., Chang, Z. et al. A two-stage image process for water level recognition via dual-attention CornerNet and CTransformer. Vis Comput (2022).
<https://doi.org/10.1007/s00371-022-02501-6>

Dou, G.; Chen, R.; Han, C.; Liu, Z.; Liu, J. Research on Water-Level Recognition Method Based on Image Processing and Convolutional Neural Networks. Water 2022, 14, 1890. <https://doi.org/10.3390/w14121890>

Contributions



Rodrigo Morales Aguayo

Separated data into different categories to create multiple linear regression models with adjustments to the variables.

Documented changes made.



Andrés Olvera Rodríguez

Researched various ML models to improve the linear regression models with new additions, and documented new changes.



Carlos Estrada Ceballos

Created CNN with images divided into years and seasons, including segmentation and image preprocessing.

Documented changes made.



Jessica Nicole Copado Leal

Added features to the models by researching other strategies to work with the variables to obtain various results. Documented changes made.

Thank you!