**Methodology**

**Dataset**

**Data Augmentation**

Image augmentation is a key preprocessing step that makes deep learning models more reliable and useful in more situations. When training a DL model, augmentation methods are used to make the dataset more diverse and improve the model's ability to find trends in different situations. Bridge pictures are flipped horizontally and vertically, scaled, and rotated as part of the augmentation process, which adds more variety to the collection.

* **Horizontal and Vertical Flipping:** The first step in the enlargement process is to flip the original bridge images horizontally and vertically. This is done by making horizontal and vertical mirror copies of the source photos. By doing this, the model sees pictures of bridges from different angles, which helps it learn to recognize buildings no matter how they are arranged. This method is especially useful when bridges are facing different directions or when the dataset doesn't have enough different image views.
* **Scaling and Rotation:** Scaling and rotation are two more methods for enhancement that add variety like real-world situations. Scaling is the process of changing the size of a picture, either up or down, while keeping its aspect ratio. This makes it look like the camera is at different distances from the bridge, giving the impression of different points of view. spin adds a controlled amount of spin to the picture, making it look like the observer's point of view has changed. Together, these changes give the model a better understanding of how bridges look from different directions, which is important for making accurate predictions based on data that hasn't been seen yet.

By combining these methods, the bridge image dataset becomes more complete and reflects the many different ways bridges are photographed. This improved collection, which is full of different visual cues, makes it possible to train the segmentation model. In a way, the augmentation process works as a bridge between the real world's inherent complexity and the model's ability to correctly and consistently understand and segment bridge structures.

**UNET Configurations**

The UNET design is a new idea that has changed the way image segmentation is done. Its layers are set up in a complicated way, and skip connections are used in a clever way. This design has a unique symmetry structure that makes it easy to see the differences between things in images.

The two most important parts of the UNET design are the contracted path and the expanding path. The contracted path, which is also called the encoder, is made up of several convolutional layers and max-pooling processes. This part does a good job of capturing important traits of different sizes. As the picture moves along the contracting path, its size gets smaller and feature maps are made that store increasingly generalized information.

On the other hand, the encoder, which is part of the large path, is made up of convolutional layers that have been turned around. During this step, the picture is carefully brought back to its original spatial detail. Incorporation of feature maps from the contracting path through skip links is the real innovation. These links allow low-level, fine-grained data to be sent to the broad path, which makes it easier to find the exact location of segmented structures.

Most of the time, the contracted path is made up of three blocks. Each block has two convolutional layers and then a max-pooling layer. These blocks gradually make the network's receptive field bigger, which lets it pick up information about its surroundings. At the same time, the structure of the expanding path is the same as that of the shrinking path, but the twists are in the opposite order, which gradually up samples the feature maps. By making skip links between the layers of the contracting and expanding lines, feature maps can be put together to bring together low-level and high-level details.

In a nutshell, the UNET design combines multi-scale contextual cues, which makes it great for difficult image segmentation projects. The model is able to solve spatial problems by putting layers in a symmetrical pattern and using skip links in a smart way. This is made possible by a deep understanding of how images are built. Through this thorough investigation, we hope to figure out the complicated mechanisms that make the UNET design the best, especially when it comes to segmenting bridges.

**Model Training**

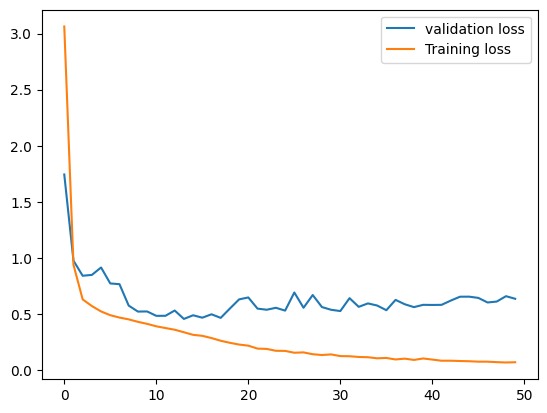
Several crucial procedures are conducted during the bridge segmentation model's training to guarantee its correctness and efficacy. The training set and the test set are originally separated from the collection of bridge photos. The training set, which serves as the model's training data, receives around 77% of the total data. The test set, which makes up the remaining 33%, is used to validate and evaluate the performance of the trained model. The UNET model, which is renowned for its ability to reliably detect objects inside photos, has been selected as the architecture for setting the model. For the architecture to better portray the delicate features of bridge structures, layers and skip connections are arranged symmetrically.

The sparse categorical cross-entropy loss function is used, and the RMSprop optimizer is used to perform optimisation during training. By assessing the discrepancy between the anticipated segmentation and the actual ground-truth masks, this loss function is optimal for the job of segmenting images. A learning rate of 0.001 is used to govern how quickly the model responds to the data, ensuring efficient convergence. A batch size of 8 pictures is further specified, enabling effective computing without compromising gradient accuracy. A whole loop of the training dataset is represented by each of the model's 100 epochs of training. The test set is periodically used to assess the model's performance throughout the training phase. When compared to the ground-truth annotations, metrics like Intersection over Union (IoU) and pixel accuracy may be used to determine how well the model is able to detect bridge structures.

The UNET model is rigorously fine-tuned to precisely recognize and outline bridge structures inside pictures by systematically carrying out these stages. With this thorough methodology, the model is not only properly designed but also equipped to handle the complexity of real-world settings, producing accurate and dependable bridge segmentation results.

**Results**

The UNET model's effectiveness on both the initial and enhanced datasets is highlighted by the outcomes of its training and evaluation. Initial training and validation losses for the model were 0.01 and 0.75, respectively, when it was just trained on the initial data used for training without any augmentation. Figure 1 shows the development of all of these losses throughout the course of training.



After completing the first training phase, test samples were used to evaluate the model's effectiveness by visualizing its output. Figure 2 shows the segmentation results for a variety of test samples, illustrative of the model's capacity to distinguish bridge structures.

A collage of images of bridges

Description automatically generated

The sets used for training and testing both included enhanced samples to improve the model's generalizability. On this expanded dataset, the UNET model then experienced a second training session. The model obtained training and validation losses of 0.13 and 0.21 at this phase, respectively. Figure 3 shows how these losses changed during the course of the increased training.

A graph showing the loss of a training

Description automatically generated

After this enhanced training, the model's effectiveness was assessed by using test samples. Figure 4 displays the bridge image segmentations that were produced using the enhanced trained UNET model. These graphics show how the model can segment bridge structures precisely even when given additional data.

A collage of images

Description automatically generated

The study of the data indicates significant tendencies in the UNET model's performance across various training conditions. The UNET model's initial training produced a significantly smaller training loss of 0.01 utilizing only the original samples. The model trained with enhanced samples, on the other hand, showed a larger training loss of 0.13. Surprisingly, the validation loss for the supplemented model, which was 0.21, was lower than the validation loss for the baseline model. The difference between the training and validation losses reveals a significant difference between the two models. The second model performs better on bridge photos that haven't been encountered before because of its smaller validation loss, which signals a more effective generalization to unseen data.

The behavior of the model is further clarified by examining the loss curves. The first model's loss curves show a clear distinction between training and validation losses as well as significant variations in validation loss. Such differences frequently point to overfitting, when the model absorbs noise from the training set and performs poorly on fresh data. The loss curves of the second model, on the other hand, provide a more favourable scenario. The training and validation loss curves are close together, indicating a balanced convergence, and the validation loss converges without experiencing any notable changes. This result highlights the durability and efficiency of the second model, demonstrating its ability to consistently perform on both training and unknown samples.

Overall, the findings highlight the value of augmentation in enhancing the model's generalization skills. The upgraded model performed better on unknown data, as seen by its reduced validation loss, whereas the baseline model had lower training loss. The improved model's effectiveness and stability are highlighted by the loss curve comparison in addition to the convergence of the training and validation curves. Together, these revelations highlight the augmented model's increased potential in practical situations, reiterating the need of diverse training data for reliable model performance.