Data Augmentation Strategy for RASMD

Overview

This report describes the data augmentation pipeline used for experiments on the RASMD dataset, which contains paired RGB and SWIR images for object detection in challenging lighting and environmental conditions.

Since the dataset is multi-modal, the main focus while designing augmentations was to:

- Keep RGB and SWIR images spatially aligned
- Handle noise and variations specific to each sensor
- Ensure the augmentations work well with multi-modal detection and fusion experiments

All the augmentations discussed here are implemented and visualized in the accompanying Jupyter notebook.

Design Constraint

In RASMD, each RGB image has a corresponding SWIR image from the same scene. Because of this, spatial alignment between the two modalities must be preserved.

Any geometric transformation (such as flips or rotations) is applied in the same way to both RGB and SWIR images.
If this is not done, the images become misaligned, which negatively affects feature learning and detection performance.

Augmentation Categories

Alignment Preserving Geometric Augmentations

These augmentations are applied **jointly** to RGB and SWIR images using the same randomly sampled parameters.

The following transformations are used:

- Random horizontal and vertical flips
- Small-angle rotations
- Random crop followed by resize
- Affine transformations (translation and scaling)

Why these are used:

- To improve spatial invariance
- To handle viewpoint and scale variations
- To maintain pixel-level correspondence between RGB and SWIR

Modalit -Specific Appearance Augmentations

RGB and SWIR images have very different intensity distributions and physical characteristics. Because of this, appearance-based augmentations are applied **separately** to each modality.

RGB augmentations include:

- Brightness and contrast jitter
- Mild color variations
- Light Gaussian blur

SWIR augmentations include:

- Intensity scaling
- Adding Gaussian noise
- Contrast stretching normalization

Why these are used:

- To improve robustness to sensor-specific noise
- To reduce overfitting to clean or ideal conditions
- To encourage the model to learn more generalizable features

Translation Based Augmentation using Pix2Pix

Apart from standard augmentations, image-to-image translation using a Pix2Pix-style approach was also explored as an additional strategy to handle class imbalance in the dataset.

In RASMD, certain object classes and conditions (especially under low visibility or challenging environments) are underrepresented. Instead of relying only on oversampling or aggressive geometric augmentations, translation-based augmentation was used to generate realistic variations of existing samples.

Using paired RGB–SWIR data, the Pix2Pix-style translation framework was leveraged to:

- Translate images across conditions while preserving scene structure
- Generate visually consistent samples that resemble underrepresented cases
- Increase diversity without breaking spatial alignment

The translated images maintain object layout and semantics, making them suitable for training object detection models.

Why Translation Helps with Class Imbalance

Traditional augmentation techniques mainly modify appearance or geometry but do not introduce new visual distributions. In contrast, translation-based augmentation:

- Creates realistic domain variations
- Helps the model see rare conditions more frequently
- Reduces bias toward dominant classes and environments

This makes it a useful complementary strategy, especially when collecting additional real data is not feasible.

Design Considerations

- Translation was used as a data enrichment strategy, not a replacement for real samples
- Generated images were visually inspected to ensure realism
- Alignment between RGB and SWIR modalities was preserved

Augmentations Explicitly Avoided

Some common augmentations were intentionally not used:

- CutMix and MixUp (they break RGB–SWIR alignment)
- Independent random crops for each modality
- Aggressive random erasing

These were avoided to prevent spatial inconsistency between paired images.

Relation to Model Experiments

The augmentation pipeline was designed to support:

- Modality-aware feature learning
- Experiments with gating and feature-level fusion
- Robust learning when one modality is noisier than the other

Although full multi-modal fusion experiments were limited due to dataset constraints, the augmentation strategy itself is consistent and reusable for similar paired multi-sensor datasets.

Summary

Overall, the augmentation strategy combines alignment-preserving geometric transforms, modality-specific appearance augmentations, and translation-based data generation. Together, these approaches improve robustness, help mitigate class imbalance, and support multi-modal object detection experiments on paired RGB–SWIR data