ICDEC 2024 Challenge on VDVWC: Methodologies and Novelty report

Authors: Kirti Raj (Sikkim Manipal Institute of Technology, Majitar), Ashu Priya (Vellore Institute of Technology, Chennai)

Introduction:- The objective of this project was to develop an object detection model using the AVD dataset, which consists of 2600 images with highly imbalanced class distributions. The challenge was to accurately detect various vehicle types despite the significant class imbalance and low-quality images which were captured in diverse weather conditions.

Dataset Analysis:- The dataset exhibited the following class distribution:

- **car:** 10146
- **bike**: 2379
- **auto**: 1020
- **rickshaw**: 1195
- **cycle van**: 45
- **cycle**: 291
- **taxi**: 315
- **bus**: 387
- **truck**: 331
- **van**: 336
- **minitruck**: 343
- **boat**: 166
- **motorvan**: 9
- **toto**: 132
- **train**: 2

Initial Attempts and Challenges:- Initial attempts using YOLOv5 [3] and YOLOv8 [4] models resulted in poor performance, with many classes being misclassified as background. This was attributed to the severe class imbalance and the models' inability to handle underrepresented classes effectively. Applying oversampling and agressive augmentations to under-represented classes didn't improve the model's metrices either.

"Qn. Can I train a neural network using synthetic images?

Ans. No. Or, to be more precise, you'll probably end up with a neural network that is great at detecting your synthetic images, but unable to detect much in real-world images." [2]

Methodology:- To address the issues, the RetinaNet model from FAIR's Detectron 2 [5] was selected for its robustness in handling class imbalance through the use of the **focal loss** algorithm. The focal loss adjusts the learning process to pay more attention to hard-to-classify examples, thereby improving the detection of underrepresented classes.

Another key factor contributing to object detection in low resolution images is **FPN(Feature Pyramid Network)**. It creates an architecture with rich semantics at all levels as it combines low-resolution semantically strong features with high-resolution semantically weak features. This is achieved by creating a top-down pathway with lateral connections to bottom-up convolutional layers.

Training and Data Augmentation:- Several data augmentation techniques were applied to enhance the model's performance, including:

- **Random Brightness**: Adjusting brightness with a factor range of 0.8 to 1.2.
- **Random Contrast**: Adjusting contrast with a factor range of 0.8 to 1.2.
- **Random Lighting**: Applying random lighting adjustments with a probability of 0.7.
- **Random Rotation**: Rotating images within an angle range of -10 to 10 degrees.
- **Random Flip**: Horizontally flipping images with a probability of 0.5.
- **Resize Shortest Edge**: Resizing the shortest edge of the images to one of the values in the range of 640 to 800 pixels, with a maximum size of 1333 pixels.

These augmentations helped in creating a more diverse training set, reducing the model's tendency to overfit and improving generalization.

Results:- After training the RetinaNet model with the augmented dataset, significant improvements were observed. The confusion matrix indicated a substantial reduction in the misclassification of objects as background, and better detection rates for previously underrepresented classes.

Metrics - Precision: 0.5964; Recall: 0.5982; F1 Score: 0.5970; mAP: 33.5560; mAP50: 64.6025; mAP75: 29.9240

Novelty and Contributions:- The novelty of this approach lies in the effective application of the focal loss algorithm to manage class imbalance, [1]FPN to address object detection challenges faced with low resolution features,coupled with targeted data augmentation strategies to improve detection accuracy. The use of RetinaNet provided a more balanced detection performance across all classes, addressing the key challenges posed by the dataset.

Conclusion:- The combination of RetinaNet and custom data augmentations thus proved effective in handling the imbalanced AVD dataset with images under diverse weather conditions, resulting in improved detection accuracy across all classes. This methodology can be extended to similar problems where class imbalance and low-quality images are significant challenges.

References:-

[1] https://developers.arcgis.com/python/guide/how-retinanet-works/

[2] https://www.ccoderun.ca/programming/darknet_faq/#synthetic_images

[3] https://github.com/ultralytics/yolov5/releases

[4] https://github.com/ultralytics/ultralytics

[5] https://github.com/facebookresearch/detectron2

[6] Focal Loss for Dense Object Detection, *Tsung-Yi, Priya, Ross Kaiming, Piotr, Facebok AI research (FAIR),* arXiv:1708.02002v2 [cs.CV] Wed, 7 Feb 2018 18:44:44 UTC