Data Preprocessing For Cityscapes Dataset

Introduction

Cityscapes is an easily scalable dataset targeting the recognition of semantic units in urban settings. Examples of high-resolution street views with tight pixel-based annotations cut across multiple instance types, such as automobiles, humans, streets, and buildings. The dataset turned out of utmost importance when it comes to the optics and application of deep learning in such sectors like autonomous driving and smart cities, across the globe. The training of these deep learning models calls it for a challenge due to the complex nature hence handling proper preprocessing is recommended.

In this review, an in-depth study of how the Cityscapes datasets looks and it s appropriate natural abilities will be documented. These include image improvement by resizing, histogram equalization, and data use optimization in data augmentation among other techniques. Such efforts address the most significant issues with the deep learning workflow and are integral in making the model more accurate and less biased.

Dataset and Methodology

1. Data Collection And Schema

Cityscapes data set has two important and inseparable parts- the training and inference sets each of them has many images.

leftImg8bit_trainvaltest.zip

a) contains 5000 Polygonal 8-void -order Training, Validation, and Testing images — pictures of the urban environment in various cities, retain the same split among those subsets of pictures.

b) Each image is 8-bit and has a typical size of around 1024x2048 pixels. These high resolution images are very critical for segmentation purposes.

gtFine_trainvaltest.zip

a) consists of fine annotated every pixels of an image in the dataset in terms of road, pets, cars, sidewalks, and others.
b) For each image in the dataset there are several annotation files:
   - _color.png — visual representation of the image with its annotation for better understanding,
   - instance_ids.png — encodes objects (e.g. cars, people) instances by unique ids,
   - label ids.png — a file where every pixel is assigned object classes e.g. car or road,
   - polylines are in the file polygons.json — the detailed shapes of the objects contained in the image to give the actual outlines.

2. How Large And Detailed Is The Dataset

There are accompanying advantages and obstacles related to the volume and quality of the image data:

On the one hand: The detailed annotations and image definitions are intricate enough to be used for segmenting several objects such as people or road signs accurately and precisely.

On the other hand: With the images being of high quality (1024x2048 pixels), the large size of each makes the memory and computational resources used during training much higher, requiring pre-processing in order to efficiently utilize the resources.

Methodology

To be able to train a model in deep learning, there is a need to change all the original data and make it evolving form, as follows: thus, the preprocessing of the dataset is necessary:

1. Image resizing

The Cityscapes dataset contains pictures of high quality. These images require a lot of time for processing by the systems without particularly large RAM. Resizing shrinks the image dimensions by maintaining the valuable scene features.

- Scope: The initial photographs (1024x2048) were downsized to 512x1024 to provide the requisite details and reduce the computational load.
- Algorithm: The resizing application facilitated the smooth transformation of the images using the bilinear interpolation while the annotation files i.e. segmentation masks were resized to maintain the pixel value and class borders which is known as nearest point interpolation.

2. Normalizing the images.

The aim of normalization is to standardize the pixel values of images to being less chaotic and more user-friendly to the neural network that is processing them. This standardization aids in the optimization of the model's convergence by making sure the pixel values is on the same intensity levels.

a) Range: Instead of using the original pixel values, the pixel values were later normalized within the range of 0 and 1 such that the raw pixel values were divided by 255.
b) Mean and Standard Deviation: There could also be different values for the different channels as seen in models like ResNet or VGG:

   - Mean: [0.485, 0.456, 0.406], which corresponds to the R, G, and B channels in order.
   - Standard Deviation: [0.229, 0.224, 0.225].

Models that are typically pre-trained on ImageNet would particularly benefit from this kind of normalisation as it corresponds to the poissons input distribution of the model.

3. Data Augmentation and Image Processing

Data augmentation is how the number of the training data is enhanced without directly impacting the magnitude of the data. It further aids better generalization by fitting several extra data disturbance factors that resemble real world, every day, in practice use cases, for example, different perspectives such as angles or even instances of changing light.

- Horizontal Flipping: It goes without saying that the horizontal line that should be drawn in the dead center of the given image is quite important. This format is less preferred due to the effective training of the imbalance data, At Bottom: 50% probability was used for horizontal flipping as street scenes are to some extent symmetric.
- Random Cropping: Random cropping is necessary while training, e.g. (512x512) in order to zoom into various parts of an image and enable the model to deal with the smaller objects in the image every time.
- Color Jitter: Variables such as brightness, contrast and saturation illuminate changes in lighting which could be such situations in the world.

4. Preprocessing the Annotations

.The regression targets, or segmentation masks, have also undergone a rescale operation using nearest neighbor interpolation method for an purposeful edit. The augs such as rotation, zooming and scaling were applied uniformly to the images as well as to the masks allowing the masks to align the labels with the masks.

Future Objectives

a) **Data Modelling:** Start the process of training an advanced level segmentation model (like U-Net or DeepLabV3). This training uses data that has already been modified specifically for Cityscapes.

b) **Model Performance:** Perform an evaluation by the model using such indicators as instance and semantic levels (IoU, and Pixel-wise Accuracy) among many others in order to estimate the quality of the segmentation that was obtained.

c) **Hyperparameter Fine-Tuning:** Examine the adjustments required in hyperparameters (e.g., learning rate, batch size) as well as the use of different model types in order to enhance model performance.

d) **Interpret Explainability:** Inspect the Explainability of AI methods are used in this study - for example LIME and SHAP - especially when it comes to specific predictions whose abnormality leads to the low confidence level.