**Case study: Lane Detection for Autonomous Vehicles using Computer Vision Algorithm**

|  |  |
| --- | --- |
| **Roll Number** | **Name** |
| CB.EN.P2AID20002 | Abhishek Gopinath |
| CB.EN.P2AID20010 | Alan Henry |
| CB.EN.P2AID20024 | Jiss Joseph Thomas |

**1. Problem Statement/Objective:**

To detect lanes on the given dataset of video or images of roads or as real-time, using computer vision algorithms which could be helpful in the proper implementation of autonomous driving.

**2. Dataset Description:**

The dataset given to our problem statement could be either as a set of images or as a video format. There are more than 1000 images for the dataset containing images and two or three videos for detecting the same.

|  |  |
| --- | --- |
| **Type** | Colour Image |
| **Size** | Multiple sizes |
| **Row \* Column** | NA |
| **Resolution** | NA |
| **Colour Model** | RGB |
| **Format** | PNG |
| **Data Acquisition** | Web, Camera |
| **No of classes** | 3 |
| **No of Images** | 853 |
| **Annotation** | PASCAL VOC format. |

**Sample Images: **

**3. Analytical Questions/Statistical Questions/Prediction level Question**

**Analysis**

1.How many vehicles in the frame

2. what are the objects detected in it

3. How many pedestrians are detected in it

4. Detecting road signs.

5. Objects on the interest area.

6. Lines inside the interest area.

7. Traffic signals detected

8. Wet area inside the interest area.

9. Dividers inside the frame.

10. How far the lanes to be detected.

**Analytics**

1. Why vehicles are less in a specific area

2. Why more pedestrians are found at a certain point

3. Why the road is wet?

4. What type of road we are traveling on.

5. What time do most people drive

6. Do people dim their light while passing

7. How often a certain route is taken.

8. How many speed breakers were detected to analyze whether there is a school nearby.

9. Why there is no horn sign detected

10. Why vehicle in the front is slowing down

**4. Block Diagram:**

****

**5. Preprocessing:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Domain** | **Filter** | **Input Image** | **Output Image** |
| **Spatial Domain** | **Image Smoothing**  ***Average Filter*** |  |  |
| ***Weighted Average Filter*** |  |  |
| ***Gaussian Blurring*** |  |  |
| ***Median Filter*** |  |  |
| ***Image Sharpening*** |  |  |
| ***Roberts Filter*** |  |  |
| ***Sobel Filter*** |  |  |
| ***Gamma Transform*** |  |  |
| ***Log Transform*** |  | gamma = 0.8 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Domain** | **Filter** | **Input Image**  **(with added noise)** | **Output Image** |
| **Frequency Domain** | ***Low Pass Filter*** | **Decentralized Image:** |  |
| ***High Pass Filter*** | **Decentralized Image:** |  |
| ***Ideal Low Pass Filter*** |  |  |
| ***Ideal High pass Filter*** |  |  |
| ***Butterworth Low Pass Filter*** |  |  |
| ***Butterworth High Pass Filter*** |  |  |
| ***Gaussian Low Pass Filter*** |  |  |
| ***Gaussian High Pass Filter*** |  |  |

**Observation after Comparison Between Spatial and Frequency Domain Filters:**

**Spatial Domain:**

**Input -> Image Processing -> Output**

**Frequency Domain:**

**Frequency + Distribution -> Image Processing -> Inverse Transformation -> Output**

* The spatial domain deals with the image plane itself whereas the Frequency domain deals with the rate of pixel change.
* The spatial domain works based on direct manipulation of pixels whereas the Frequency domain works based on modifying Fourier Transform.
* The spatial domain takes less time to computer whereas the Frequency domain takes more time to compute.

**6. List of Features:**

|  |  |  |
| --- | --- | --- |
| **Feature Name** | **Purpose** | **Category**  **[Image/Vision]** |
| Edges | To detect lanes | Image or Video |
| Blobs | To detect vehicles | Image or Video |
| Corner | To identify the type of vehicles o object, present in the data | Image or Video |
| Ridges | To identify the shape and outline the objects | Image or Video |

**7. Feature Detection and Tracking:**

* ***Harris Corner Detection:***

1.Take the grayscale of the original image

2. Apply a Gaussian filter to smooth out any noise

3. Apply Sobel operator to find the x and y gradient values for every pixel in the grayscale image

4. For each pixel p in the grayscale image, consider a 3×3 window around it and compute the corner strength function. Call this its Harris value.

5. Find all pixels that exceed a certain threshold and are the local maxima within a certain window (to prevent redundant dupes of features)

6. For each pixel that meets the criteria in 5, compute a feature descriptor.

****

* ***SIFT******(Scale-Invariant Feature Transform):***

1. Scale-space Extrema Detection (Feature point (also called key point) detection)
   * Potential location for finding features.
2. Feature point localization
   * Accurately locating the feature key points.
3. Orientation assignment
   * Assigning orientation to key points.
4. Feature descriptor generation.
   * Describing the key points as a high dimensional vector.
5. Key point Matching



* ***SURF (Speeded-Up Robust Features):***

1. Feature Extraction
   * Integral Images
   * Hessian matrix-based interest points
   * Scale-space representation
2. Feature Description
   * Orientation Assignment
   * Extract Descriptor Components

* ***Multiscale Oriented Patches Descriptor (MOPS):***

|  |  |  |
| --- | --- | --- |
| **Multiscale Oriented Patches Descriptor (MOPS)** | ***Translation*** (***T = MT1)*** |  |
| ***Rotation*** (***T = MRMT1)*** |  |
| ***Affine Transformation*** |  |
| ***Perspective Transformation*** |  |
| ***Scaling*** |  |

#### **After Feature Detection and Tracking**:



**8. Scene:**

|  |  |  |
| --- | --- | --- |
| **Scene** | **Feature to be detected** | **Sample Image** |
| **Image of some cyclists on road** | Faces (Pedestrian or Cyclists) |  |
| **Image of a road with lanes visible** | Lanes |  |
| **Image of a road with lanes and rarely visible cars** | Cars |  |

**9. List of objects in the scene and the features:**

|  |  |
| --- | --- |
| **Object Name** | **List of features for the object** |
| Lanes | Shape,  Color,  Texture |
| Vehicles | Shape,  Color,  Size |
| Pedestrians/Cyclists | Shape,  Color,  Size |
| Signals and Sign boards | Shape,  Color,  Size |

**10. Face Detection**

**Viola-Jones Algorithm:**







**Number of faces detected: 9**

**Number of falsely detected faces: 1**

**11. Face detection algorithms and Deep learning architecture:**

|  |  |  |  |
| --- | --- | --- | --- |
| Face Detection Algorithm | URL | Dataset | Deep learning architecture |
| Facenet | <https://machinelearningmastery.com/how-to-develop-a-face-recognition-system-using-facenet-in-keras-and-an-svm-classifier/#:~:text=FaceNet%20is%20a%20face%20recognition,of%20face%20recognition%20benchmark%20datasets.&text=About%20the%20FaceNet%20face%20recognition,implementations%20and%20pre%2Dtrained%20models>. |  | Facenet |
| Histogram of Oriented Gradients using Dlib | https://www.pyimagesearch.com/2014/11/10/histogram-oriented-gradients-object-detection/ |  |  |
| Haar Cascade Classifiers using OpenCV | https://becominghuman.ai/face-detection-using-opencv-with-haar-cascade-classifiers-941dbb25177 |  |  |
| CNN based face detector from dlib | https://towardsdatascience.com/cnn-based-face-detector-from-dlib-c3696195e01c |  | dlib |

**12. Performance Metrics:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Category** | **Purpose** | **Formula** |
| Peak Signal to Noise Ratio (PSNR) | Spatial Domain | Gives the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. | Where is the maximum possible pixel value of the image; is the Mean Squared Error; |
| Mean Squared Error (MSE) | Spatial Domain | Measures the average squared difference between the estimated values and the actual value. It is a risk function, corresponding to the expected value of the squared error loss. | Where is the Mean Squared Error; is the number of data points; are the observed values; are the predicted values;  There is no correct value. The closer it is to zero better and if its zero it means the model is perfect. |
| Structural Similarity Index (SSIM) | Spatial Domain | It is a perceptual metric that quantifies image quality degradation\* caused by processing such as data compression or by losses in data transmission. | Where is the Structural Similarity Index;  is the average of ; is the average of ; is the variance of ; is the variance of ; is the covariance of and; and are two variables to stabilize the division with a weak denominator; |
| Intersection Over Union (IoU) | Spatial Domain | It is essentially a method to quantify the percent overlap between the target mask and our prediction output. Used as object detection evaluation metrics. | Ideally, **IoU**>0.5 is considered a good prediction. |
| Accuracy |  |  |  |
| Precision |  |  |  |
| Recall |  |  |  |
| F1-score |  |  |  |

**13. Deep Learning Architectures**

[Identify the applications where the deep learning architectures has been applied]

[refer DL\_1.pdf]

[refer: <https://developer.ibm.com/technologies/artificial-intelligence/articles/cc-machine-learning-deep-learning-architectures/>]

[refer: <https://www.analyticsvidhya.com/blog/2017/08/10-advanced-deep-learning-architectures-data-scientists/>]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Architecture Name | Category | Learning | Year | Applications |
| CNN (Convolutional neural networks) |  | Supervised Learning |  | Image recognition, video analysis, and natural language processing |
| RNN (Recurrent neural networks) |  | Supervised Learning |  | Speech recognition and handwriting recognition |
| SOM (Self organizing Maps) |  | Unsupervised Learning |  | Dimensionality reduction, clustering high-dimensional inputs to 2-dimensional output, radiant grade result, and cluster visualization |
| AE(Autoencoders) |  | Unsupervised Learning |  | Dimensionality reduction, data interpolation, and data compression/decompression |
| LSTM (Long Short-Term Memory) | RNN | Supervised Learning |  | Image and video captioning systems |
| GRU (Gated Recurrent Unit) | RNN | Supervised Learning |  | Natural language text compression, handwriting recognition, speech recognition, gesture recognition, image captioning |
| RBM (Restricted Boltzmann Machines) | AE | Unsupervised Learning |  | Dimensionality reduction and collaborative filtering |
| DBN (Deep Belief  Networks) |  |  |  | Image recognition, information retrieval, natural language understanding, and failure prediction |
| DSN (Deep Stacking Networks) |  |  |  | Information retrieval and continuous speech recognition |
| LeNet | Spatial Exploitation |  | 1998 |  |
| AlexNet | Spatial Exploitation | Supervised Learning | 2012 |  |
| VGGNet | Spatial Exploitation | Supervised Learning | 2014 |  |
| GoogleNet | Spatial Exploitation | Supervised Learning | 2014 |  |
| ResNet | Depth + Multi-Path | Supervised Learning | 2015 |  |
| ResNext | Width | Supervised Learning |  |  |
| DenseNet | Multi-Path |  | 2017 |  |
| PolyNet | Width |  | 2017 |  |
| PyramidalNet | Width |  | 2017 |  |
| YOLO (You Only Look Once) |  |  |  |  |
| SqueezeNet |  |  |  |  |
| SegNet |  |  |  |  |
| GAN (Generative Adversarial Network) |  |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Architecture Name** | **Year** | **Main contribution** | **Parameters** | **Error Rate** | **Depth** | **Category** | **Reference** |
|  |  | - First popular CNN architecture | 0.060 M | [dist]MNIST: 0.8  MNIST: 0.95 | 5 |  | (LeCun et al.  1995) |
| **AlexNet** | 2012 | - Deeper and wider than the LeNet  - Uses Relu, dropout and overlap Pooling  - GPUs NVIDIA GTX 580 | 60 M | ImageNet: 16.4 | 8 |  | (Krizhevsky et al.  2012) |
| **ZfNet** | 2014 | -Visualization of intermediate layers | 60 M | ImageNet: 11.7 | 8 | Spatial Exploitation | (Zeiler and  Fergus 2013) |
| **VGG** | 2014 | * Homogenous topology * Uses small size kernels | 138 M | ImageNet: 7.3 | 19 |  | (Simonyan and Zisserman 2015) |
| **GoogLeNet** | 2015 | - Introduced block concept  - Split transform and merge idea | 4 M | ImageNet: 6.7 | 22 |  | (Szegedy et al.  2015) |
| **Inception-V3** | 2015 | * Handles the problem of a representational bottleneck * Replace large size filters with small filters | 23.6 M | ImageNet: 3.5  Multi-Crop: 3.58  Single-Crop: 5.6 | 159 | Depth + Width | (Szegedy et al.  2016b) |
| **Highway Networks** | 2015 | - Introduced an idea of Multi-path | 2.3 M | CIFAR-10: 7.76 | 19 | Depth + Multi-Path | (Srivastava et al.  2015a) |
| **Inception-V4** | 2016 | - Split transform and merge idea Uses asymmetric filters | 35 M | ImageNet: 4.01 | 70 | Depth +Width | (Szegedy et al.  2016a) |
| **Inception-**  **ResNet** | 2016 | - Uses split transform merge idea and  residual links | 55.8M | ImageNet: 3.52 | 572 | Depth + Width +  Multi-Path | (Szegedy et al.  2016a) |
| **ResNet** | 2016 | - Residual learning  - Identity mapping based skip connections | 25.6 M  1.7 M | ImageNet: 3.6  CIFAR-10: 6.43 | 152  110 |  | (He et al. 2015a) |
| **DelugeNet** | 2016 | - Allows cross layer information flow in  deep networks | 20.2 M | CIFAR-10: 3.76  CIFAR-100: 19.02 | 146 | Multi-path | (Kuen et al.  2018) |
| **FractalNet** | 2016 | - Different path lengths are interacting with each other without any residual connection | 38.6 M | CIFAR-10: 7.27  CIFAR-10+: 4.60 CIFAR-10++: 4.59  CIFAR-100: 28.20 CIFAR-100+: 22.49  CIFAR100++: 21.49 | 20  40 | Multi-Path | (Larsson et al.  2016) |
| **WideResNet** | 2016 | - Width is increased and depth is decreased | 36.5 M | CIFAR-10: 3.89  CIFAR-100: 18.85 | 28  - | Width | (Zagoruyko and Komodakis 2016) |
| **Xception** | 2017 | - Depth wise convolution followed by point  wise convolution | 22.8 M | ImageNet: 0.055 | 126 | Width | (Chollet 2017) |
| **Residual**  **Attention Neural Network** | 2017 | - Introduced an attention mechanism | 8.6 M | CIFAR-10: 3.90  CIFAR-100: 20.4  ImageNet: 4.8 | 452 | Attention | (Wang et al. 2017a) |
| **ResNeXt** | 2017 | - Cardinality  - Homogeneous topology  - Grouped convolution | 68.1 M | CIFAR-10: 3.58  CIFAR-100: 17.31  ImageNet: 4.4 | 29  - 101 | Width | (Xie et al. 2017) |
| **Squeeze & Excitation**  **Networks** | 2017 | - Models interdependencies between feature-maps | 27.5 M | ImageNet: 2.3 | 152 | Feature-Map Exploitation | (Hu et al. 2018a) |
| **DenseNet** | 2017 | - Cross-layer information flow | 25.6 M  25.6 M  15.3 M  15.3 M | CIFAR-10+: 3.46 CIFAR100+:17.18 CIFAR-10: 5.19  CIFAR-100: 19.64 | 190  190  250  250 | Multi-Path | (Huang et al.  2017) |
| **PolyNet** | 2017 | * Experimented structural diversity * Introduced Poly Inception module   - Generalizes residual unit using polynomial compositions | 92 M | ImageNet: Single:4.25 Multi:3.45 | -  - | Width | (Zhang et al.  2017) |
| **PyramidalNet** | 2017 | - Increases width gradually per unit | 116.4 M  27.0 M  27.0 M | ImageNet: 4.7  CIFAR-10: 3.48  CIFAR-100: 17.01 | 200  164  164 | Width | (Han et al. 2017) |
| **Convolutional Block Attention Module (ResNeXt101 (32x4d) +**  **CBAM)** | 2018 | - Exploits both spatial and feature-map information | 48.96 M | ImageNet: 5.59 | 101 | Attention | (Woo et al. 2018) |
| **Concurrent Spatial & Channel**  **Excitation Mechanism** | 2018 | - Spatial attention  - Feature-map attention  - Concurrent placement of spatial and channel attention | - | MALC: 0.12  Visceral: 0.09 | - | Attention | (Roy et al. 2018) |
| **Channel Boosted CNN** | 2018 | - Boosting of original channels with additional information rich generated  artificial channels | - | - | - | Channel Boosting | (Khan et al. 2018a) |
| **Competitive Squeeze & Excitation Network CMPE-**  **SE-WRN-28** | 2018 | - Residual and identity mappings both are used for rescaling the feature-map | 36.92 M  36.90 M | CIFAR-10: 3.58  CIFAR-100: 18.47 | 152  152 | Feature-Map Exploitation | (Hu et al. 2018b) |

**14. Training, Testing, and Validation**

[Place the contents that are under :]

<https://lionbridge.ai/training-data-guide/>]

[Provide your inference for 10-rule]]

12.

Performance Metrics

[refer: URL: https://towardsdatascience.com/20-popular-machine-learning-metrics-part-1-classification-regression-evaluation-metrics-1ca3e282a2ce]

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Category** | **Purpose** | **Formula** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Can also refer:

a.Refer the links , place the metric name,purpose,formula,ideal value for the metric(i,e what is the

expected valued of the metric)

Performance Metrics:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

https://towardsdatascience.com/20-popular-machine-learning-metrics-part-1-classification-regression-evaluation-metrics-1ca3e282a2ce

https://blog.floydhub.com/a-pirates-guide-to-accuracy-precision-recall-and-other-scores/

https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/

https://scikit-learn.org/stable/auto\_examples/model\_selection/plot\_precision\_recall.html

https://www.analyticsvidhya.com/blog/2020/09/precision-recall-machine-learning/

https://towardsdatascience.com/a-look-at-precision-recall-and-f1-score-36b5fd0dd3ec

13.Deep Learning Architectures

[Identify the applications where the deep learning architectures has been applied]

[refer DL\_1.pdf]

[refer : <https://developer.ibm.com/technologies/artificial-intelligence/articles/cc-machine-learning-deep-learning-architectures/>]

[refer: https://www.analyticsvidhya.com/blog/2017/08/10-advanced-deep-learning-architectures-data-scientists/]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Architecture Name | Category | Learning | Year of Design | Applications |
| LSTM | RNN | Supervised Learning |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

14.Training ,Testing and Validation

[Place the contents that is under :

<https://lionbridge.ai/training-data-guide/>]

[Provide your inference for 10-rule]

15.

**Face Detection**

**Training Dataset Description:**

**Testing Dataset Description:**

**Validation Dataset Description:**

**Search for “****Face detection with viola jones and deep learning**” ,run the code and present the results for your dataset. Understand the code related to face detection.

**16.Deep learning architecture**

a. Choose one deep learning architecture other than Convolutional Neural Network related to the application

b. Deep learning architecture Diagram

c.Working **methodology** of the architecture

**d.**

|  |  |
| --- | --- |
| **Parameter Name**  **Purpose**  **Value that is generally provided** | **Hyper Parameter Name**  **Purpose**  **Value that is generally provided** |
|  |  |
|  |  |
|  |  |
|  |  |

**References**

1. URL of the dataset

https://tryolabs.com/resources/introductory-guide-computer-vision

**References**

1. <https://github.com/overtunned/lane_detection/tree/main/Dataset>
2. [GitHub - rslim087a/road-video: Video required for finding lane lines](https://github.com/rslim087a/road-video)
3. [CULane dataset](https://xingangpan.github.io/projects/CULane.html)
4. [Real-time detection of road lane-lines for autonomous driving](https://www.researchgate.net/publication/331478663_Real-Time_Detection_of_Road_Lane-Lines_for_Autonomous_Driving)
5. [Multi-Lane Detection and Tracking Using Vision for Traffic Situation Awareness](https://ieeexplore.ieee.org/document/9253415)
6. [Real-Time Tracking and Lane Line Detection Technique for an Autonomous Ground Vehicle System](https://link.springer.com/chapter/10.1007/978-981-15-0633-8_156#:~:text=Vehicle%20tracking%20has%20been%20performed,using%20adaptive%20background%20subtraction%20technique.)