VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Artificial Intelligence

DA – 2

Part – a: Advantages And Disadvantages of HOG and spatial Pooling with Proof

Team(2):

Ramireddy Jeevan Reddy – 21BCE5163
B.Venu Vihari Yadav – 21BCE5232

The Algorithms That We Have Chosen are:

Histogram of Oriented Gradients (HOG) :

It is a feature descriptor widely used for object detection in computer vision tasks. It captures the distribution of gradient orientations in an image, providing a compact representation of its structure. HOG is particularly effective for detecting objects with distinct edges and textures, making it a popular choice for pedestrian detection, face detection, and more.

Spatial Pyramid Pooling (SPP) :

It is a technique used in computer vision for handling input images of arbitrary sizes and aspect ratios. It divides the input image into a grid of fixed-size regions and extracts features independently from each region. SPP allows neural networks to process images of different sizes while maintaining spatial information, making it particularly useful for tasks like object detection and image classification. It helps in achieving spatial invariance and improving the performance of deep learning models on images with varying sizes and aspect ratios.

ML-Model Chosen TO Implement The Image Detection Algorithms is :

- ➔ Support Vector Machine(SVM)

Advantages And Disadvantages:

Histogram Of Oriented Gradients :

Advantages:

Robustness to Illumination categorization tasks, deep learning models may outperform HOG-based methods by learning more complex and discriminative features directly from raw data, leading to improved classification accuracy and generalization features can accurately detect pedestrians under varying lighting conditions, including shadows and highlights.

Translation Invariance:
Proof: HOG captures local texture and shape information by analyzing gradient orientations, making it partially invariant to small translations or shifts in the object's position.
Example: In face detection, HOG features can detect faces at different locations and orientations within the image, despite slight variations in head position or pose.

Simplicity and Efficiency:
Proof: HOG features are computationally efficient to compute and provide a compact representation of object appearance, making them suitable for real-time applications.
Example: In video surveillance systems, HOG features can be computed quickly and used for real-time pedestrian detection on streaming video feeds.

Interpretability:

Proof: The HOG descriptor represents object appearance in terms of gradient orientations, making it interpretable and intuitive for understanding the discriminative features used by the model.
Example: In forensic analysis, HOG features can be used to identify specific objects or individuals based on their unique visual characteristics, such as edge orientations and textures.

Versatility:
Proof: HOG features can be applied to a wide range of object detection tasks, including pedestrian detection, face detection, and general object recognition.
Example: In autonomous driving systems, HOG features can be used to detect various objects on the road, including pedestrians, vehicles, and traffic signs, contributing to overall scene understanding.
Disadvantages:

Limited Robustness to Scale and Rotation:
Proof: HOG features are less robust to changes in scale and rotation of objects in the image, leading to reduced detection accuracy for objects at different sizes and orientations.
Example: In aerial imagery analysis, HOG may struggle to detect objects with varying sizes and orientations, such as buildings and vehicles, due to scale and perspective changes.

Dependency on Preprocessing:
Proof: HOG requires careful preprocessing of input images, such as resizing and normalization, to ensure consistent feature extraction across different samples.
Example: In medical imaging, HOG features may require preprocessing steps such as image registration and intensity normalization to standardize input images for accurate feature extraction.

Difficulty with Complex Backgrounds:
Proof: In cluttered or complex backgrounds, HOG may struggle to distinguish objects of interest from the background noise, leading to false positives or reduced detection accuracy.
Example: In wildlife monitoring, HOG features may produce false detections of animals in dense vegetation or camouflage, where the background texture may resemble the animal's appearance.

Lack of Semantic Information:
Proof: HOG focuses on low-level image features related to edges and textures, but it does not capture high-level semantic information about objects, limiting its performance in tasks requiring semantic understanding of the scene.
Example: In image captioning, HOG features may not provide sufficient context or semantic information to generate accurate descriptions of complex scenes or events.

Handcrafted Features:
Proof: HOG features are handcrafted and may not capture all relevant information in the data, limiting their discriminative power compared to features learned automatically from data in deep learning approaches.
Example: In fine-grained categorization tasks, deep learning models may outperform HOG-based methods by learning more complex and discriminative features directly from raw data, leading to improved classification accuracy and generalization performance.

Spatial Pyramid Pooling :

Advantages:

Handling Variable Input Sizes:

Proof: SPP was introduced to hyperparameter of handling images with variable sizes. The original paper on SPP by He et al. (2014) demonstrated its effectiveness in handling image classification tasks with input images of arbitrary sizes.

Example: In object detection tasks, objects may appear at different scales and aspect ratios. SPP allows the model to capture spatial information effectively, regardless of the input image's size, leading to improved detection accuracy.

Preservation of Spatial Information:

Proof: SPP divides the input image into a spatial pyramid structure and computes features at multiple scales. This strategy enables the model to capture fine-grained spatial details and contextual information.

Example: In pedestrian detection, SPP can capture features at different resolutions, allowing the model to detect pedestrians at various scales and orientations accurately.

Improved Robustness to Scale and Translation:

Proof: SPP's multi-scale pooling strategy helps in capturing features at different resolutions, making the model more robust to scale variations and translations.

Example: In face detection, SPP can detect faces of different sizes and orientations in complex scenes with varying lighting conditions and background clutter.

Adaptability to Complex Scenes:

Proof: SPP's hierarchical feature representation enables the model to capture spatial information at multiple levels of abstraction, making it more adaptable to complex scenes with cluttered backgrounds and occlusions.
Example: In scene understanding tasks, SPP can effectively distinguish between different objects of interest and background clutter, leading to more accurate object localization.

Reduced Dependency on Pre-processing:

Proof: Unlike traditional feature extraction methods like HOG, SPP can handle input images of varying sizes directly, reducing the need for pr-processing steps such as resizing and normalization.
Example: In image classification, SPP can process images of arbitrary sizes without loss of spatial information, simplifying the overall pipeline and improving efficiency.

Disadvantages:

Computational Complexity:

Proof: SPP involves computing features at multiple scales and pooling regions, leading to increased computational complexity compared to simple feature extraction methods like HOG.

Example: In real-time applications such as autonomous driving, the increased computational overhead of SPP may impact system performance and response time.

Model Complexity and Overfitting:

Proof: The introduction of additional parameters and complexity into the model with SPP may increase the risk of overfitting, especially when trained on small datasets.

Example: In fine-grained object recognition tasks, SPP may learn to memorize training samples rather than generalize to unseen examples, resulting in reduced generalization performance.

Difficulty in Interpretation:

Proof: SPP's hierarchical feature representation and multi-scale pooling strategy make it challenging to interpret and visualize compared to simpler feature descriptors like HOG.

Example: In medical image analysis, understanding the learned representations of SPP may require domain expertise and advanced visualization techniques to extract meaningful insights.

Data Requirements:

Proof: SPP may require larger amounts of training data compared to simpler feature extraction methods like HOG to learn discriminative features effectively, especially for complex scenes and object classes.

Example: In fine-grained categorization tasks such as bird species recognition, SPP may require a diverse dataset with a large number of samples to capture subtle differences between classes accurately.

Hyperparameter Tuning:

Proof: SPP introduces additional hyperparameters related to the spatial pyramid configuration (e.g., pyramid levels, pooling sizes) that need to be tuned carefully to achieve optimal performance.

Example: In image segmentation tasks, selecting appropriate hyperparameters for SPP, such as the number of pyramid levels and pooling sizes, can significantly impact segmentation accuracy and computational efficiency.

<u>Comparison Of Two Algos With Respect To Accuracy Scores :</u>

- ➔ Histogram Of Gradients:

```
In [162]: y_pred = svm.predict(X_test)
          accuracy = accuracy_score(y_test, y_pred)
          print("Accuracy:", accuracy)

          Accuracy: 0.2893401015228426
```

- ➔ Spatial Pyramid Pooling

```
In [14]: accuracy = accuracy_score(y_test, y_pred)
         print("Accuracy:", accuracy)

         Accuracy: 0.5253807106598984
```

<u>Conclusion:</u>

<u>**HOG Accuracy:**</u> The model trained using HOG features achieved an accuracy of approximately 28.93%. While HOG features are robust to illumination changes and offer simplicity and efficiency in feature extraction, they may struggle with complex backgrounds and have limited robustness to scale and rotation changes. As a result, the model's performance may be hindered, especially on datasets with diverse object scales and orientations.

<u>**SPP Accuracy:**</u> On the other hand, the model trained using Spatial Pyramid Pooling (SPP) features achieved a significantly higher accuracy of approximately 52.54%. SPP addresses some of the limitations of HOG by preserving spatial information at multiple scales and offering improved robustness to scale and translation changes. It handles variable input sizes effectively and adapts well to complex scenes with cluttered backgrounds. As a result, the SPP-based model demonstrates superior performance compared to the HOG-based model.

Conclusion:

In conclusion, while both HOG and SPP are popular feature extraction techniques for object detection, SPP offers significant advantages over HOG in terms of accuracy and robustness, as demonstrated by the higher accuracy achieved by the SPP-based model. SPP's ability to handle variable input sizes, preserve spatial information, and adapt to complex scenes makes it a preferred choice for object detection tasks, particularly in scenarios with diverse object scales, orientations, and backgrounds. Therefore, when selecting a feature extraction method for object detection, considering the specific requirements of the task and dataset, SPP may offer better performance and more reliable results compared to HOG.

~ Thank You ~