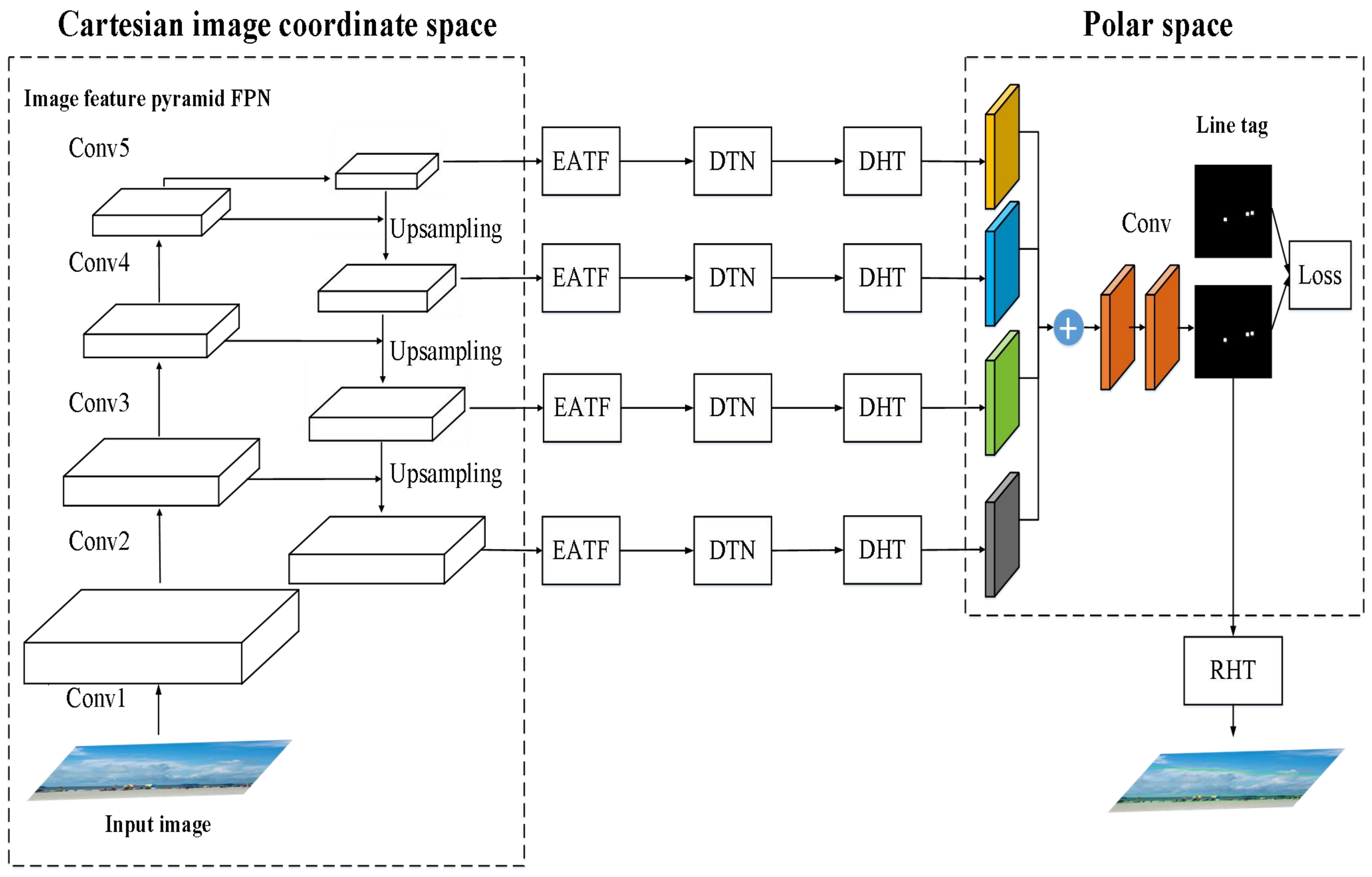
Distortion Correction Sub-Network for Semantic Segmentation based on Deep Hough Transform

Wanpeng Geng, Jing Liu, Dexin Zhang, and Hui Zhang

Tianjin University of Science & Technology, China

Abstract

This study proposes a novel method for semantic line correction utilizing deep Hough transform, aimed at tackling the challenges associated with detecting semantic straight lines and correcting image distortions in natural scenes. Traditional approaches frequently consider semantic straight line detection as a subset of object detection or simply adapt conventional object detection techniques, thereby neglecting the inherent characteristics of straight lines and consequently leading to suboptimal performance. Herein, we employ a deep Hough transform-based algorithm to achieve semantic line detection in images. The adopted approach utilizes parameterization and the Hough transform to map depth representations into parameter space for straight line detection, effectively exploiting the geometric properties of lines. Innovatively, we introduce the Distortion Correction Sub-network (DTN) to mitigate image distortion and enhance the success rate of deep Hough transform line detection. Furthermore, the DTN can dynamically adjust its spatial transformation according to various image transformations, thereby achieving effective image distortion correction. Experimental results demonstrate that the proposed method outperforms previous state-of-the-art methods on both self-constructed and publicly available datasets, thus substantiating its efficacy and superiority in addressing the challenges of semantic line detection and image distortion correction.



Deep Hough transform agent semantic secant detection network

Methods

In this study,the Deep Hough Transform (DHT) converts Cartesian coordinates of the image into polar coordinates, while the Reverse Hough Transform (RHT) performs the reverse transformation, yielding the results of semantic segmentation line detection. The DHT module in this study is based on previous work, with key differences in the coordinate transformation process: the image center is set as the origin, the polar angle spans from 0 to 2π, and a consistent sampling rate is used. The maximum polar radius varies across different feature layers in the image pyramid, with multiple polar radii $[\nabla r_1, \nabla r_2, \ldots\ldots, \nabla r_n]$ designed to align feature layers with polar coordinate space layers of uniform dimensions, improving computational efficiency.

The Edge Activation Function Network (EATF) is introduced to improve semantic line detection by leveraging edge information critical for distinguishing boundaries in complex scenes. Unlike traditional edge detection methods that rely on binary images from the Hough transform, EATF targets boundary-focused segmentation lines. It achieves this through two main components: a channel self-attention mechanism, inspired by SENet, to adaptively weight convolutions and prioritize useful features, and a Tanh activation function that reduces sensitivity to sharp edges, helping the model focus on prominent semantic lines. Additionally, using the ReLU function in channel attention prevents gradient issues, enhancing training efficiency.

The segmentation task leverages a dual loss function: $L = L_{line} + L_{cls}$ minimizes cross-entropy between predictions and a Gaussian-smoothed annotation map, and L_{line} uses multi-class cross-entropy for line categorization. These adjustments streamline the training process and enhance line localization accuracy in polar coordinate space.

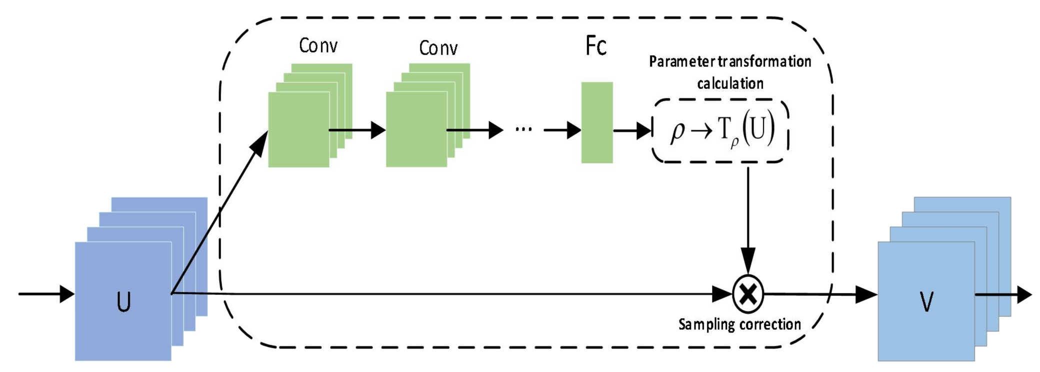
Methods

The Dynamic Transformation Network (DTN) facilitates dynamic spatial transformations on images, allowing for a range of transformation types. It is a modular component independent of the neural network (NN), making it adaptable for integration at various stages, such as pre-object detection, post-image processing, or after convolutional feature mapping. This flexibility enables its application in diverse tasks like classification, segmentation, and image distortion correction.

The DTN module incorporates several components: a spatial transformer, a channel self-attention mechanism, matrix transformations, sampling correction, and the Edge Activation Function (EATF) module. The primary focus in terms of parameters is the channel self-attention mechanism and fully connected layers, with the attention mechanism's parameters typically one-sixteenth of the input channel count.

Computationally, the DTN module's demands mainly stem from convolutional operations and fully connected layers. These demands are influenced by the size of the input feature maps and the number of parameters involved. Despite the high parameter count and computational load, the integration of the self-attention mechanism and EATF module enhances the network's expressiveness and performance, particularly in distortion correction tasks.

DTN training is deeply integrated into the overall network training pipeline, with its primary goal being to enhance image fidelity, particularly in object detection tasks. It achieves this improvement by correcting distortions without the need for extra annotations specifically for distorted images, which makes it highly efficient. As a modular component, the DTN is designed to integrate seamlessly into the end-to-end training process. This allows for significant improvements in detection accuracy and robustness, all while preserving the efficiency of the training flow without introducing any disruptions.



Distortion Correction Sub-Network (DTN)

Results

The table shows the results of multiple methods on SEL and NKL datasets, focusing on F-measurement, accuracy, and recall metrics. For SEL, the results from HED+Hough and SLNet are referenced, while other methods use ResNet50 as the backbone. Our proposed method, especially the one that incorporates the EATF module, achieves the highest f measurement on the threshold, a significant improvement over SLNet and hough based methods.

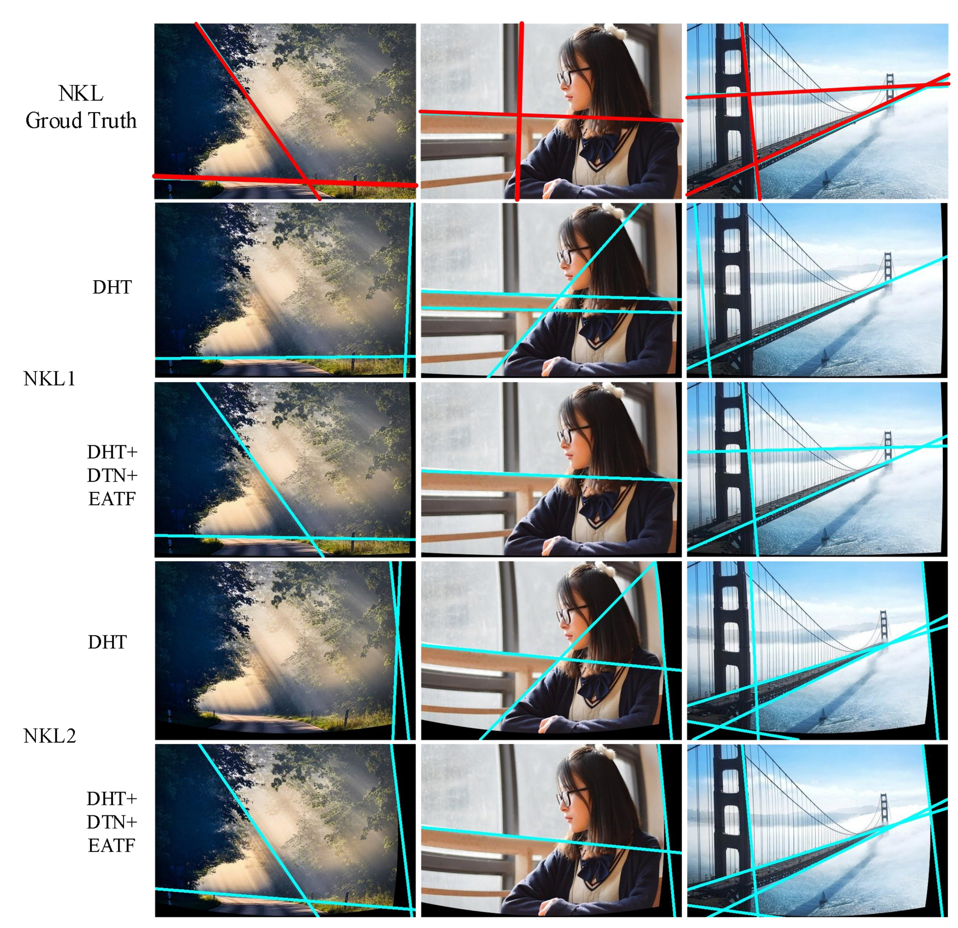
In the test on the NKL dataset, SLNet's training code is not available, and comparisons are made between Hough, DHT+DTN, and DHT+DTN+EATF. The DHT+DTN+EATF method, as shown in the NKL results in the table, is consistent with ground truth values even when the data set is distorted.

Based on the NKL1 and NKL2 data results, our approach outperforms the DHT approach in terms of accuracy, recall, and F-measure, highlighting the greater ability to reconstruct target details from the original data set, despite distortions. In addition, since the NKL2 dataset exhibits more pronounced distortions than NKL1, the detection efficiency decreases, confirming the hypothesis that the greater the degree of distortion in the dataset leads to an increase in detection complexity and a corresponding decrease in performance.

Dataset	Method	P（%）	R（%）	F（%）
SEL	HED+Hough	35.60	42.00	38.50
	SLNet	76.20	72.90	74.50
	DHT	**82.86**	74.52	78.47
	DHT+DTN	77.76	80.37	79.04
	DHT+DTN+EATF	78.27	**81.69**	**79.95**
NKL	HED+Hough	21.30	62.20	31.80
	DHT	68.42	76.65	72.30
	DHT+DTN	70.16	80.12	74.81
	DHT+DTN+EATF	**72.92**	**77.43**	**75.10**
NKL1	DHT	62.37	72.75	67.16
	DHT+DTN+EATF	**67.63**	**74.47**	**70.89**
NKL2	DHT	29.09	58.32	38.82
	DHT+DTN+EATF	**37.27**	**63.12**	**46.86**

Experimental results and data comparison

Examples of detection results of different methods on SEL dataset



Comparison of different methods in NKL distorted dataset

Results

Both SLNet and HT methods use HED edge detectors and NMS for image preparation, and SLNet also involves iterative fine-tuning, which can slow down reasoning. In contrast, our method requires only one forward pass and a simplified NMS to get faster results, as shown in the figure on the left on the SEL dataset.In distortion testing of NKL datasets, our method outperforms DHT, and despite the distortion, still aligns better with Ground Truth, especially at image edges. The figure on the right shows the superior performance of our method in processing distorted data.

Conclusion

This paper introduces the integration of the deep Hough transform with distortion correction subnetworks for the general task of detecting semantic segmentation lines. The proposed distortion correction subnetwork fully leverages deep learning and classical Hough transform to establish a robust global context, while allowing the network to dynamically adapt its feature extraction process to various image transformations. Extensive experimental results unequivocally demonstrate that the fusion of deep Hough transform with DTN yields superior performance and exceptional generalization capability when juxtaposed against extant methodologies. Furthermore, the proposed method demonstrates a trade-off between accuracy and efficiency. In future work, we will incorporate large language models and use CLIP to enhance the description of semantic segmentation lines, generating the required semantic segmentation lines.