



VG-SSL: Benchmarking Self-supervised Representation Learning Approaches for Visual Geo-localization

Jiuhong Xiao, Gao Zhu, Giuseppe Loianno

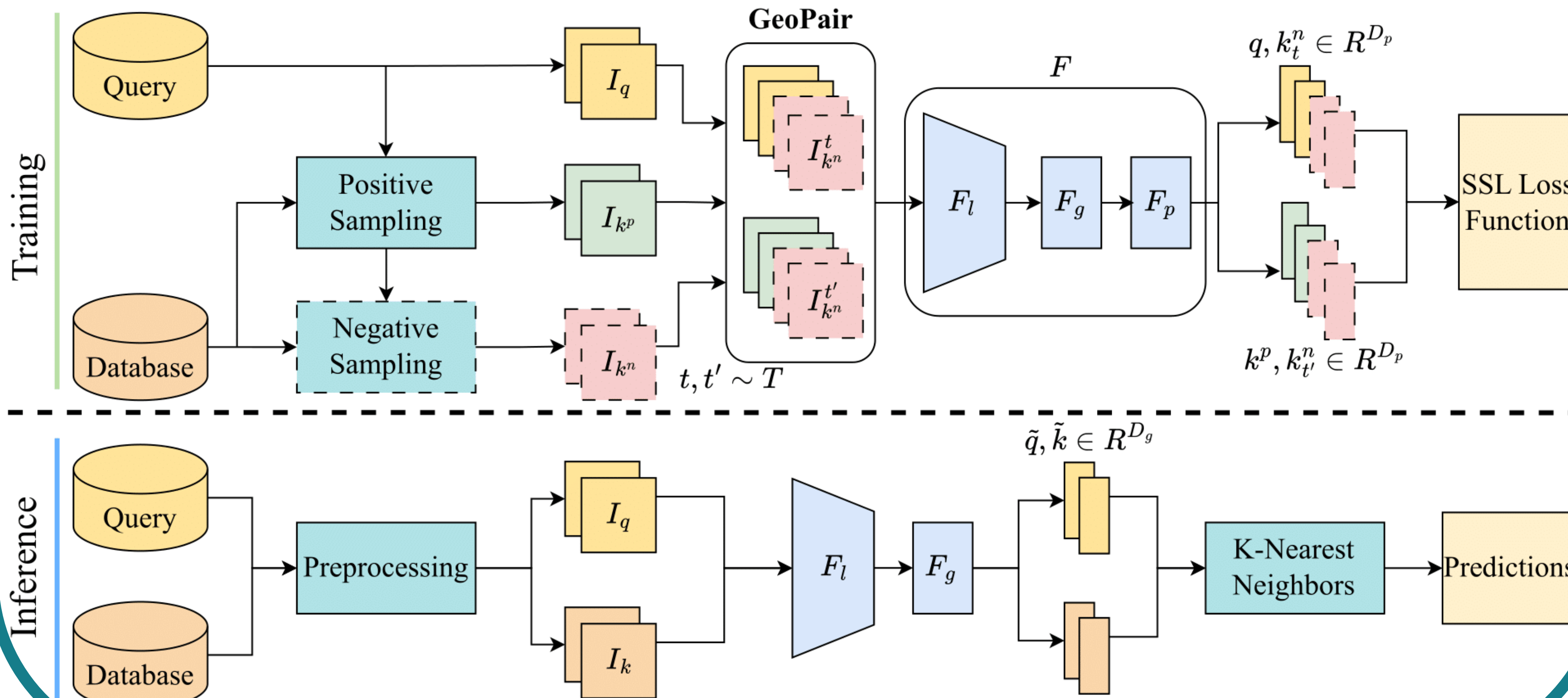
Agile Robotics and Perception Lab, Tandon School of Engineering, New York University, USA

Introduction

This work introduces the **VG-SSL** framework, a versatile platform for integrating and benchmarking diverse self-supervised learning (SSL) methods in representation learning for visual geo-localization (VG).

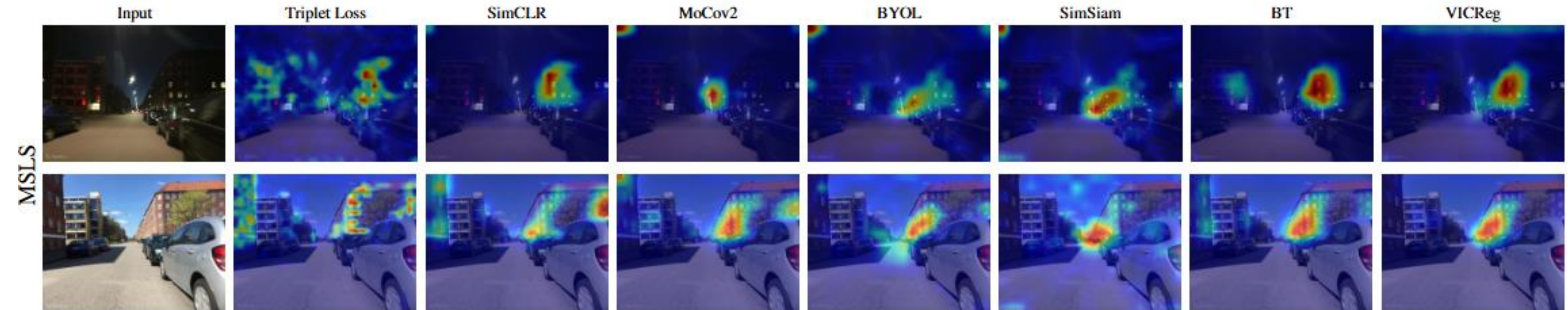
Method Overview

- **GeoPair**: A geo-related pairing mechanism for effective representation learning.
- **Data Augmentation**: Constructs negative samples to avoid mode collapse in SSL.
- **Projection Layers**: Utilized in training but omitted in inference.



Main Results

	D_g	MSLS Val			MSLS Challenge			Pitts30k			Tokyo24/7			Nordland		
		R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10
<i>Our One-Stage Methods with ResNet50-GeM</i>																
Triplet Loss (Baseline) [8]	1024	76.9	86.1	89.5	53.5	68.1	72.3	76.7	89.1	92.3	50.2	67.9	76.8	39.5	59.0	67.7
SimCLR	1024	84.2	92.2	94.2	63.1	78.9	83.6	82.8	91.9	94.6	54.6	74.9	81.9	39.9	56.4	63.9
MoCov2	1024	81.5	90.5	92.8	59.0	73.8	79.2	82.6	92.4	95.1	51.4	68.3	76.5	28.0	42.7	50.1
BYOL	1024	72.7	85.5	87.7	50.4	66.4	71.4	80.2	91.5	94.4	44.8	63.8	70.8	10.6	18.5	23.5
SimSiam	1024	75.0	85.8	88.6	52.1	67.0	72.2	78.6	89.8	92.7	51.1	67.6	71.4	12.5	21.5	27.0
Barlow Twins	1024	79.5	89.5	91.9	59.2	74.2	79.1	80.8	91.7	94.2	45.7	61.9	70.8	18.5	30.5	38.0
VICReg	1024	77.4	89.3	91.2	58.0	74.1	79.0	80.2	91.3	94.1	50.2	65.4	74.3	14.9	25.1	31.3
<i>Our One-Stage Methods with DeiT-S</i>																
Triplet Loss (Baseline) [56]	256	79.3	90.5	92.7	54.9	75.1	79.6	72.9	88.5	92.6	43.5	65.7	72.4	21.4	33.7	41.0
SimCLR	256	81.1	91.1	93.1	58.9	77.1	82.6	84.7	93.9	96.0	59.4	76.2	80.0	24.9	38.9	46.1
MoCov2	256	76.1	88.5	91.1	56.8	75.2	78.7	80.8	92.4	95.0	50.8	69.8	77.1	15.4	26.4	33.0
BYOL	256	58.2	75.3	79.6	37.7	54.0	60.4	76.6	89.4	92.9	43.2	62.2	68.6	4.1	7.9	10.6
SimSiam	256	56.2	76.2	80.1	35.3	52.3	58.7	79.7	91.0	93.6	47.3	63.8	74.0	6.2	11.5	15.4
Barlow Twins	256	79.7	91.4	93.1	59.1	76.1	81.5	82.6	92.1	95.0	58.4	75.2	80.6	28.1	43.3	51.1
VICReg	256	75.8	89.5	91.9	56.9	74.0	78.2	81.7	92.3	95.2	51.7	66.7	74.6	19.3	32.1	39.6



Key Takeaways

- **Performance**: Contrastive Learning (SimCLR, MoCov2) and Information Maximization (Barlow Twins, VICReg) methods outperform Self-distillation (BYOL, SimSiam) in VG.
- **Comparison to Triplet Loss**: Selected SSL methods show overall better performance and produce more condensed heatmaps due to differing objective functions.