Geometry-Structured Perception for Topology-Agnostic Robotic Rebar Tying

Huiguang Wang[1], Zekai Jin[1], and Yi Shao[1,*]

[1]Department of Civil Engineering, McGill University, Montreal, Quebec, Canada
[*]Corresponding Author
huiguang.wang@mail.mcgill.ca, zekai.jin@mail.mcgill.ca, yi.shao2@mcgill.ca

Abstract –
Robust perception of rebar joints is challenging due to geometric variability, irregular intersection topologies, and real-world sensing imperfections. Existing methods largely rely on appearance-driven recognition and topology-specific supervision, which limits generalization. We reformulate rebar joint perception as a geometry-dominated structural understanding problem and propose a two-stage learning pipeline that biases learning toward geometric invariants without explicit domain adaptation. In the first stage, a geometry-only synthetic dataset is constructed to train a detector that is used exclusively as a fixed annotator to generate reliable pseudo-labels for simple cross-shaped intersections in real images. In the second stage, a final perception model is trained from scratch using only these pseudo-labeled, background-removed real images, introducing authentic geometric variability without manual annotation. Based on this formulation, we further design a unified, topology-agnostic perception-to-action pipeline for robotic rebar tying. Experiments demonstrate robust sim-to-real generalization across diverse rebar topologies and configurations.

Keywords –
Geometry-dominated perception, Rebar joint detection, Sim-to-real generalization, Topology-agnostic perception, Robotic rebar tying

1 Introduction

Reliable perception of rebar joints remains a central challenge for robotic rebar tying in real-world environments. Most existing vision-based approaches formulate this problem as appearance-driven recognition and rely on manually annotated datasets [1-4], where joint-specific bounding boxes or keypoints are labeled. Such methods heavily depend on low-level visual cues and are therefore sensitive to variations in texture [5] and illumination [6], leading to limited robustness under realistic sensing conditions.

Beyond appearance variability, rebar meshes exhibit diverse and irregular geometric intersection patterns, including orthogonal, non-orthogonal, multi-bar clusters, and irregular joints with tie bars. Current perception models [1-4] are topology-specific and require collecting new annotated datasets when encountering previously unseen intersection configurations, which severely limits scalability and generalization. Although synthetic datasets can partially reduce annotation cost, significant sim-to-real domain gaps persist as a result of heterogeneous camera modalities and real-world sensing noise that is difficult to replicate in synthetic images [7].

A large body of prior work addresses sim-to-real transfer through domain adaptation [8] and domain randomization [9]. Domain adaptation methods typically aim to reduce the gap between synthetic and real domains by aligning feature distributions via adversarial learning [10] or style transfer [11], while domain randomization increases visual diversity by aggressively varying textures, lighting, and rendering parameters in simulation. While these approaches are effective at mitigating appearance-level discrepancies, they often operate under the assumption that improved visual alignment is sufficient for downstream task generalization. In the context of rebar joint perception, however, appearance variation is often not the dominant source of failure. Instead, performance degradation is primarily driven by geometric variability, irregular intersection topologies, and real-world structural imperfections that are difficult to capture through appearance-focused adaptation or randomization alone. Consequently, visual domain alignment alone is insufficient to address the core challenges of rebar perception.

These observations suggest that rebar joint perception should be formulated to prioritize geometric structure while progressively weakening reliance on appearance-specific cues, rather than treating it as a pure appearance recognition or domain alignment problem. Instead of explicitly matching visual domains, we seek to bias the learned representation toward geometric invariants that remain stable across sensing conditions and intersection topologies.

The main contributions of this work are summarized as follows:

(1) Geometry-dominated formulation and representation for rebar joint perception: We reformulate rebar joint perception as a geometry-dominated structural understanding problem, where

keypoints defined and annotated in simulation. By prioritizing geometric invariants, our approach avoids explicit modeling of appearance factors such as textures, rib patterns, and lighting, thereby reducing modeling complexity and supervision dimensionality. As a result, effective learning can be achieved with significantly smaller datasets while weakening reliance on appearance-specific cues.

(2) Two-stage sim-to-real learning via geometry-dominated supervision and offline pseudo-labeling: We propose a two-stage learning pipeline in which strong geometric priors are first established using texture-free synthetic data. A geometry-only model trained based on synthetic data is then fixed and used exclusively to generate pseudo labels for structurally simple intersections in real images. These pseudo-labeled samples [12] are subsequently used in a separate training process, introducing geometric imperfections from real-world sensing.

(3) Unified topology-agnostic perception-to-action pipeline for rebar tying: We design a unified tying gun action formulation that directly input 2D optimal tying angles. Then, a 3D collision-free optimization is applied to get the final collision-free tying pose, enabling consistent and geometrically feasible tying pose generation across diverse rebar configurations without topology-specific templates or task-dependent action definitions.

2 Method

2.1 Two-stage Sim-to-real Learning via Geometry-dominated Supervision and Offline Pseudo-labeling

To bias learning toward geometric structure while weakening reliance on appearance-specific cues, we adopt a two-stage training strategy. The first stage is used solely to generate reliable geometric pseudo-labels, while the second stage performs the actual training of the final perception model. No manual annotation is used in either stage.

2.1.1 Stage One: Geometry-Only Synthetic Training for Pseudo Label Generation

To generate reliable geometry-only supervision without manual annotation, we construct a fully automatic synthetic dataset in the PyBullet simulation environment (Figure 1). Rather than pursuing photorealistic rendering, the dataset is intentionally designed to contain only geometric information, ensuring that the learned detector focuses on structural patterns of rebar intersections instead of appearance variations.

We only simulate two simple intersection typologies: orthogonal and non-orthogonal cross-shaped joints. Although simple, these configurations capture the essential geometric characteristics of rebar intersections. Orthogonal joints represent canonical right-angle crossings, while non-orthogonal joints introduce angular variation and asymmetry, encouraging the model to learn geometry-invariant cues. All rebars are modeled as smooth, texture-free round bars, eliminating texture and material cues. For each simulated scene, PyBullet provides precise analytic geometry, from which the intersection keypoint is automatically derived. The rendered output consists of background-removed RGB images preserving only rebar silhouettes.

Figure 1. Synthetic dataset generation in PyBullet

Data acquisition follows a controlled geometric design. We generate intersections at five angles (70°, 75°, 80°, 85°, and 90°), vary camera viewpoints across four positions, and sample two camera-to-joint distances (400 mm and 500 mm). In addition, each configuration is rotated in-plane, with images rendered every 5°, resulting in a total of 2,880 synthetic samples. Then, a geometry-aware transformer, shown in Section 2.1.3, is trained on synthetic dataset solely for the purpose of pseudo label generation. The trained model reliably localizes orthogonal and non-orthogonal cross-shaped joints but generalizes poorly to more complex configurations, reflecting the limited structural diversity of the synthetic data. This limitation motivates the second training stage.

2.1.2 Stage Two: Offline Pseudo-labelling on Background-removed Real-world Images

To incorporate real-world geometric variability without manual annotation, we apply an offline pseudo-labeling procedure to real images, as illustrated in Figure 2. Background-removed real-world images are processed by a detector trained solely on synthetic data to automatically identify orthogonal and non-orthogonal cross-shaped intersections.

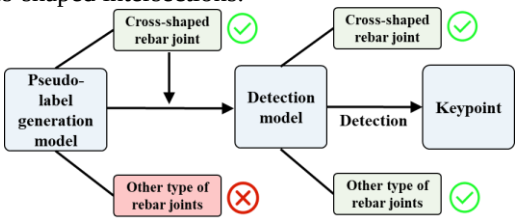


Figure 2. Two stage offline pseudo labelling

The detector's predictions on these real images are treated as pseudo-labels and used to construct a new training dataset. Importantly, this process does not reuse

synthetic images or pretrained weights from the first stage. Instead, the final perception model is trained from scratch using only the pseudo-labeled real data generated offline by the frozen geometry-only detector.

By restricting pseudo-label generation to reliable cross-shaped intersections, the proposed procedure mitigates error accumulation while introducing authentic geometric imperfections present in real-world sensing, including sensor noise, slight deformation, and imaging artifacts.

2.1.3 Network Architecture (Geometry-Aware Transformer)

To implement the proposed geometry-dominated perception formulation, we adapt an existing transformer-based keypoint estimation framework, DETRPose [13], as a practical realization rather than a methodological contribution. The original DETRPose architecture is designed for human pose estimation and predicts a fixed set of semantically distinct keypoints with a predefined skeletal topology. Such assumptions are incompatible with rebar meshes, where intersection topology is irregular and the number of joints varies across scenes.

We therefore simplify the original architecture into an end-to-end single-point detector by removing topology-specific constraints, shown in Figure 3. Specifically, we retain the backbone and encoder to extract multi-scale geometric features, while reconfiguring the decoder and prediction heads to predict a single geometry-based keypoint per query. The original multi-keypoint prediction head is replaced with a streamlined regression head that outputs only a 2D keypoint location corresponding to the center of a rebar intersection.

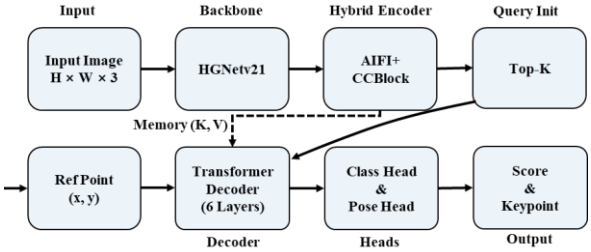


Figure 3. Geometry-aware transformer detection network

2.2 Topology-Agnostic Tying Action Formulation

2.2.1 Direction Vector Estimation

After detecting a rebar joint keypoint, the keypoint is back-projected to 3D space using the depth map to define the tying position. A local plane is then fitted to neighboring points, and its normal is used as the rebar tying direction vector shown in Figure 4. The rebar tying direction vector encodes the nominal tying direction at the joint.

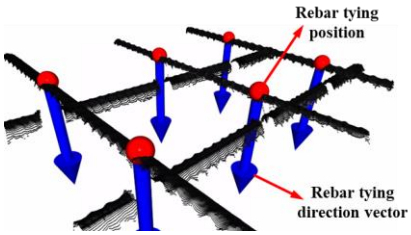
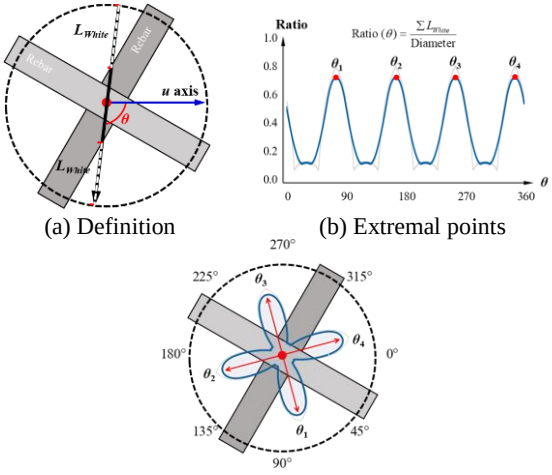


Figure 4. The tying direction vectors of rebar joints

2.2.2 The Definition of the Optimal Tying Angle Function

Given the predicted rebar joint keypoint and an associated rebar tying direction vector, an optimal tying angle function is defined to compute multiple 2D optimal tying angles. As shown in Figure 5(a), since the perception input is a background-removed image with a uniform white background, the detected joint region is also represented against a white background. Leveraging this property, we simplify the local workspace analysis by centering a circle of diameter D at the detected keypoint. For a given rotation angle, the diameter of this circle intersects both white (free space) and non-white (occupied) regions corresponding to the projected rebar geometry.



(a) Definition (b) Extremal points

(c) Visualization of action angles

Figure 5. The definition of the 2D optimal tying angle function

We define an objective function as Eq. 1 that maximizes the length of the diameter segment lying in the white region, denoted as L_{White}. By searching for extremal points of this objective over rotation angles, we obtain a set of 2D optimal rebar tying angles that maximize the available free workspace around the joint. These candidate angles correspond to orientations where the tying gun is most likely to approach the joint without interference, as illustrated in Figure 5(b) and Figure 5(c).

In practice, the objective defined in Eq. 1 may produce noisy responses over the rotation angle due to boundary discretization in the background-removed image. We therefore apply a one-dimensional smoothing operation defined in Eq. 2 to the objective curve before extremum search, and select candidate tying angles from the extrema of the smoothed objective.

As shown in Figure 6, the tying gun end-effector is initialized by aligning it with the rebar tying direction vector, and is then rotated around this direction according to the 2D optimal rebar tying angles.

$$\text{Ratio}(\theta) = \frac{\sum L_{White}}{\text{Diameter}} \qquad \textbf{Eq. 1}$$

$$\overline{\text{Ratio}(\theta)} = \frac{1}{47}\sum_{\theta-23^{\circ}}^{\theta+23^{\circ}} \text{Ratio}(\theta) \qquad \textbf{Eq. 2}$$

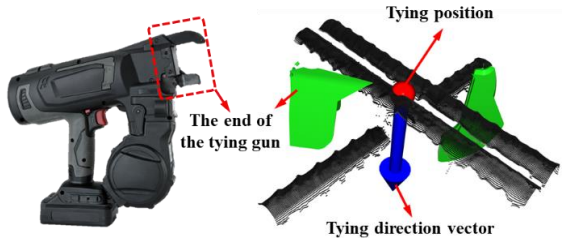


Figure 6. Tying gun alignment with direction vector and 2d optimal angles

However, due to perspective projection, the 2D optimal rebar tying angle may deviate from the true collision-free direction in 3D space. As a result, directly applying the 2D optimal rebar tying angle is not the final applied solution. To efficiently address this issue, a collision-free optimization strategy is adopted.

As illustrated in Figure 7, the 2D optimal rebar tying angle is treated as an initial estimate. A local angular search is then performed within a ±10° range around this estimate, resulting in a 20° candidate interval. For each 2D optimal rebar tying angle within this interval, the tying gun pose is evaluated using geometric collision distance computation. The angle that yields the maximum collision distance is selected as the final optimal tying angle, ensuring maximal geometric clearance under the modeled constraints.

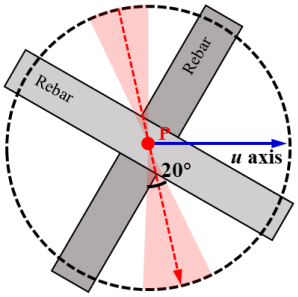


Figure 7. Collision-free optimization

3 Experiment

3.1 Experimental Setup

We evaluate the proposed geometry-dominated perception framework on real-world rebar scenes captured by three depth cameras (Figure 8): Cam-1 Mech-Mind NANO and Cam-2 NANO ULTRA structured-light cameras, and a Cam-3 Intel RealSense D435i RGB-D camera, covering different sensing accuracies and noise characteristics. Background removal is performed as a preprocessing step following the procedure described in [1]. All experiments are conducted on a workstation with an Intel i7-14700F CPU and an NVIDIA GeForce RTX 4070 GPU.

The training process follows a two-stage learning pipeline. In the first stage, a geometry-only detector is trained on a fully synthetic dataset comprising 2,880 texture-free samples generated in simulation. This detector is trained exclusively for offline pseudo-label generation and remains fixed thereafter. In the second stage, the trained detector is applied to 800 background-removed real images to generate pseudo-labels for structurally simple cross-shaped intersections. These pseudo-labeled real images are then used to train the final perception model from scratch, without reusing synthetic data or pretrained weights from the first stage.

(a) Cam-1 (c) Cam-2 (c) Cam-3

Figure 8. Adopted three types of depth sensing devices

3.2 Generalization to Seen and Unseen Topologies

The proposed geometry-dominated perception framework is designed to operate across diverse rebar intersection configurations without relying on topology-specific supervision. To comprehensively evaluate its generalization capability, we consider both seen and unseen rebar joint topologies. The seen topologies include orthogonal (Type 1) and non-orthogonal cross-shaped intersections (Type 2), which are used during synthetic training and pseudo-label generation. The unseen topologies consist of more complex configurations, such as multi-bar clusters (Type 3) and irregular joints with tie bars (Type 4) that are not represented during training.

Table 1 summarizes the Detection Success Rate (DSR) of the proposed geometry-dominated perception framework across rebar joint configurations that are either seen or unseen during training. As described in the experimental setup, orthogonal (Type 1) and non-orthogonal cross-shaped intersections (Type 2) are included during synthetic training and pseudo-label

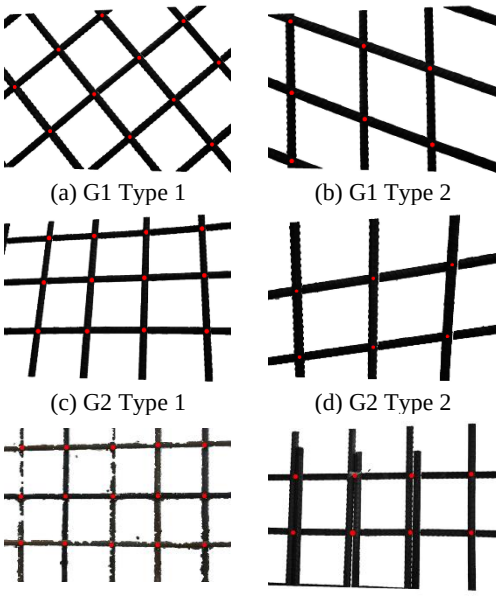
generation, while multi-bar clusters (Type 3) and irregular configurations (Type 4) are not observed during training and are treated as unseen topologies. Moreover, the detection visualizations are reported in Figure 9.

DSR is used to evaluate whether the predicted rebar joints correspond to valid and actionable tying locations. Currently, there is no publicly available dataset with ground-truth annotations for robotic rebar tying tasks. As a result, the evaluation was performed through manual inspection. Each predicted rebar joint was visually verified to determine whether it corresponds to a real rebar joint and whether the resulting tying orientation is geometrically feasible. The DSR is defined as the ratio of valid detections to the total number of evaluated predictions.

Table 1. Detection Success Rate (DSR) on Seen and Unseen Rebar Topologies

No.	Seen in Training	Joint type	Camera type	Number of images	DSR (%)
G1	Yes	Type 1 / 2	Cam-1	400	99.82
G2	Yes	Type 1 / 2	Cam-1	400	99.80
G3	No	Type 1	Cam-3	109	93.61
G4	No	Type 3	Cam-1	147	98.72
G5	No	Type 3	Cam-3	143	93.66
G6	No	Type 4	Cam-2	210	100

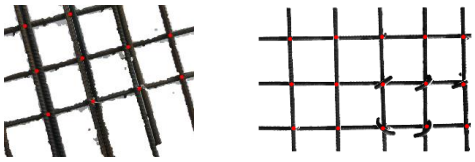
Notes: "G" means "Group". G2 is captured from a lateral viewpoint, while all other groups are captured under a vertical view. To evaluate generalization without manual annotation, we use Detection Success Rate (DSR) instead of mAP. A detection is considered successful if the predicted keypoint corresponds to a valid and actionable rebar intersection, enabling topology-agnostic evaluation.



(a) G1 Type 1 (b) G1 Type 2

(c) G2 Type 1 (d) G2 Type 2

(e) G3 Type 1 (f) G4 Type 3



(g) G5 Type 3 (h) G6 Type 4

Figure 9. Detection Results on Seen and Unseen Rebar Topologies

For the seen topologies, the proposed method achieves a detection success rate of over 99% under both vertical (G1) and lateral (G2) viewpoints. This result indicates that the geometry-aware perception framework maintains stable performance despite viewpoint changes when the underlying joint topology is familiar.

More importantly, strong generalization performance is observed on unseen topologies (G3–G6). For multi-bar cluster intersections (Type 3), the proposed method achieves DSR values of 98.72% (G4) and 93.66 % (G5), despite these configurations not being represented during training. For irregular joints with tie bars (Type 4), the DSR further increases to 100% (G6). These results indicate that the learned geometry-dominated representation generalizes effectively beyond the intersection structures encountered during training, without relying on topology-specific supervision.

It is also noteworthy that Cam-3 corresponds to a color RGB-D camera, whereas all training data consist of monochrome images. Nevertheless, comparable performance is achieved on Cam-3 for unseen configurations. In particular, the method attains DSR values of 93.61% (G3) and 93.66% (G5) on Cam-3. This observation suggests that the learned representation is largely insensitive to color information and supports the effectiveness of the geometry-dominated training strategy in reducing reliance on appearance-specific cues.

Overall, the quantitative DSR results reported in Table 1, together with the qualitative detection visualizations shown in Figure 9, demonstrate that the proposed approach achieves robust generalization across seen and unseen rebar joint topologies, different viewpoints, and heterogeneous sensing modalities, without requiring topology-specific supervision or appearance-level domain adaptation.

As summarized in Table 2, the proposed geometry-dominated approach substantially reduces modeling complexity compared to domain adaptation and domain randomization methods. By avoiding explicit modeling of appearance factors such as textures and illumination, as well as topology-specific representations, the method eliminates the need to handle the combinatorial complexity arising from diverse rebar surface patterns and intersection types. As a result, effective generalization can be achieved with significantly lower data requirements, making the proposed approach more

practical for real-world rebar joint perception.

Table 2. Comparison with Domain Adaptation (DA) and Domain Randomization (DR)

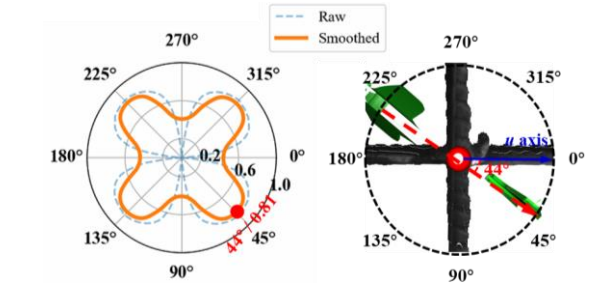
Aspect	DA	DR	Ours
Core idea	Align feature domains	Randomize appearance	**Prioritize geometry**
Main variation handled	Appearance gap	Appearance diversity	**Geometric variability**
Use of real images	For domain alignment	Optional	**Offline pseudo-labeling**
Training scheme	Joint/adversarial	Single-stage	**Two-stage, independent**
Appearance modeling	Explicit alignment	Heavy randomization	**Not required**
Modeling topology	Yes	Yes	**No**
Data requirement	Medium	High	**Low**
Suitability	Limited	Limited	**High**

3.3 Tying Action Feasibility Analysis

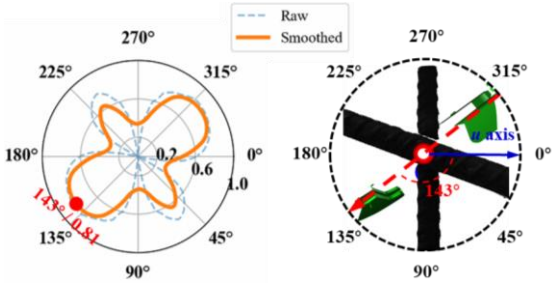
The 2D optimal rebar tying angles are shown in Figure 10. As summarized in Table 3, the proposed collision-free optimization consistently produces positive final minimum collision distances for all evaluated rebar configurations. For cross-shaped intersections (R-G1 and R-G2), the optimized poses achieve minimum clearances of 8.12 mm and 6.91 mm, respectively.

Similar feasibility is observed for previously unseen and structurally more complicated arrangements. In dense multi-bar cluster intersections (R-G4), the method maintains a minimum clearance of 4.27 mm despite the limited space created by parallel rebars. For irregular joints that include tie bars (R-G6), the final minimum distance reaches 7.60 mm, further confirming that the optimized poses remain collision-free even in complex structural configurations.

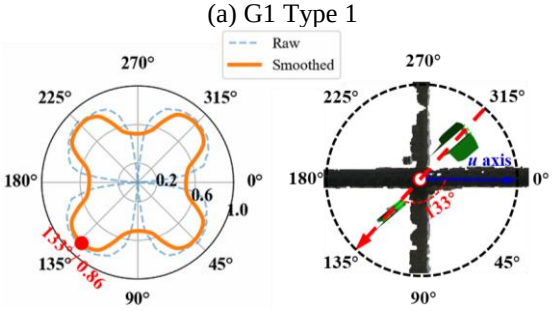
In summary, the final minimum collision distances listed in Table 3 demonstrate that the proposed collision-free optimization consistently generates feasible tying poses across a variety of rebar topologies and sensing viewpoints. Moreover, the effective tying workspace surrounding the optimized pose can be approximated as at least twice the reported minimum distance. This corresponds to a clearance range of roughly 0.5–1.5 times the diameter of standard 10 mm rebars, which is generally sufficient for practical tying tasks. Such residual workspace provides adequate tolerance for tool positioning as well as minor execution deviations in real-world construction environments.



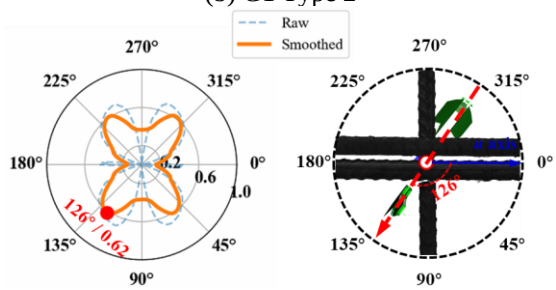
(a) G1 Type 1



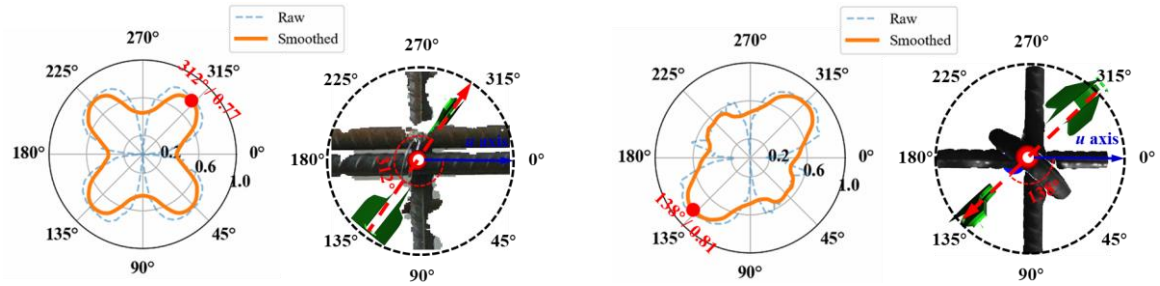
(b) G1 Type 2



(c) G3 Type 1



(d) G4 Type 3



(e) G5 Type 3 (f) G6 Type 4

Figure 10. 2D optimal rebar tying angles and loaded tying gun

Table 3. The experimental setup of rebar tying

No.	Joint type	Camera type	Average initial collision distance (2D optimal rebar tying angle)	Average final minimum distance (after collision-free optimization)
R-G1	Type 1 / 2	Cam-1	5.08 mm	8.12 mm
R-G2	Type 1 / 2	Cam-1	4.45 mm	6.91 mm
R-G3	Type 1	Cam-3	2.82 mm	4.40 mm
R-G4	Type 3	Cam-1	2.26 mm	4.27 mm
R-G5	Type 3	Cam-3	1.79 mm	3.11 mm
R-G6	Type 4	Cam-2	4.92 mm	7.60 mm

4 Limitations and Future Work

The proposed pipeline assumes that foreground rebar structures can be separated from the background using standard background removal techniques. This preprocessing step simplifies the subsequent geometric analysis by isolating the structural elements of interest.

However, the performance of this stage may degrade in scenarios where background segmentation is inaccurate. For example, highly cluttered construction environments, strong illumination variations, or partial occlusions may introduce segmentation errors that affect the quality of the reconstructed geometric representation.

Background removal is a common preprocessing step in our pipeline and can be implemented using a variety of existing depth-based or learning-based segmentation approaches. In future work, we plan to investigate integrated perception frameworks that jointly perform foreground segmentation and structural geometry detection, which may further improve robustness under complex real-world conditions.

The current study focuses on geometry-dominated perception and action feasibility evaluation and does not include physical tying execution using a robotic manipulator. However, the proposed tying action formulation is evaluated through quantitative collision distance metrics.

In future work, we plan to integrate the proposed perception and action modules with a robotic rebar tying system to evaluate execution-level performance under real construction conditions. This includes assessing robustness to actuation errors, dynamic disturbances, and tool–environment interaction. Additionally, extending the framework to handle more complex joint configurations and incorporating feedback during execution are promising directions for further investigation.

5 Conclusion

This paper presents a geometry-dominated perception and unified rebar tying action framework for robotic rebar tying. By reformulating rebar joint perception as a geometry-dominated structural understanding problem, the proposed approach weakens the reliance on appearance cues and topology-specific supervision. A two-stage learning pipeline is adopted, in which geometry-only synthetic data are used to train a fixed detector for reliable pseudo-label generation, and the final perception model is subsequently trained from scratch on real images without manual annotation. Experimental results demonstrate robust generalization across seen and unseen rebar topologies, different viewpoints, and heterogeneous sensing modalities.

Building upon the geometry-dominated perception output, a unified formulation for tying actions is introduced to generate feasible tying poses for various rebar joint configurations. The geometric validity of each

pose is quantitatively assessed using collision-free optimization, enabling systematic evaluation and optimization under realistic spatial constraints. Experiments conducted on real-world datasets show that the proposed framework demonstrates strong generalization across diverse rebar topologies, sensing conditions, and visual disturbances, while preserving high data efficiency.

Overall, this work demonstrates that emphasizing geometric structure provides an effective and scalable alternative to appearance-driven methods for rebar perception and action generation, and offers a solid foundation for future deployment in robotic rebar tying systems.

Acknowledgment

The authors gratefully appreciate the financial support provided by McGill University through McGill Engineering Doctoral Award (MEDA).

Data availability

The dataset will be made publicly available at: https://huiguangwang.top/tutorial/Topology-Agnostic-Robotic-Rebar-Tying/

References

[1] L. Deng, S. Wang, J. Guo, R. Cao, M. Liu. 3D keypoint detection-based automated rebar spacing inspection: Application for robotic integration. Advanced Engineering Informatics, 66 103418, 2025.

[2] H. Duan, M. Yu, T. Ai, M. Zhu, H. Jiang, S. Guo. YOLO-FAS: A lightweight model for detecting rebar intersections location and tying status. Neurocomputing, 624 129485, 2025.

[3] R. Feng, Y. Jia, T. Wang, H. Gan. Research on the system design and target recognition method of the rebar-tying robot. Buildings, 14 (3):838, 2024.

[4] M. Liu, J. Guo, L. Deng, S. Wang, H. Wang. Enhanced vision-based 6-DoF pose estimation for robotic rebar tying. Automation in Construction, 171 105999, 2025.

[5] K. Xu, H. Chen, C. Xu, Y. Jin, C. Zhu. Structure-texture aware network for low-light image enhancement. IEEE Transactions on Circuits Systems for Video Technology, 32 (8):4983-4996, 2022.

[6] P. Tang, X. Wang, S. Bai, W. Shen, X. Bai, W. Liu, A. Yuille. Pcl: Proposal cluster learning for weakly supervised object detection. IEEE transactions on pattern analysis machine intelligence, 42 (1):176-191, 2018.

[7] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, P. Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS), pages 23-30, Vancouver, Canada, 2017.

[8] T.-W. Huang, Y.-H. Chen, J.J. Lin, C.-S. Chen. Deep learning without human labeling for on-site rebar instance segmentation using synthetic BIM data and domain adaptation. Automation in Construction, 171 105953, 2025.

[9] T. Sun, B. Han, S. Rusinkiewicz, Y. Shao. Rebar grasp detection using a synthetic model generator and domain randomization. Automation in Construction, 176 106252, 2025.

[10] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, V. Lempitsky. Domain-adversarial training of neural networks. Journal of machine learning research, 17 (59):1-35, 2016.

[11] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, pages 2223-2232, Venice, Italy, 2017.

[12] D.-H. Lee. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Workshop on challenges in representation learning, ICML, pages 896, Atlanta, USA, 2013.

[13] S. Janampa, M. Pattichis. DETRPose: Real-time end-to-end transformer model for multi-person pose estimation. arXiv preprint, 2025.