Fusión de imágenes de profundidad obtenidas con sistemas LiDAR y de Estereovisión por medio de técnicas de aprendizaje profundo

Autores: Miguel Angel Molina Garzón - Henry Dario Mantilla Claro

Director: Hoover Rueda-Chacón

Escuela de Ingeniería de Sistemas
Universidad Industrial de Santander
Bucaramanga, Colombia

AGENDA

01 Introducción

02 Objetivos

03 Método Propuesto

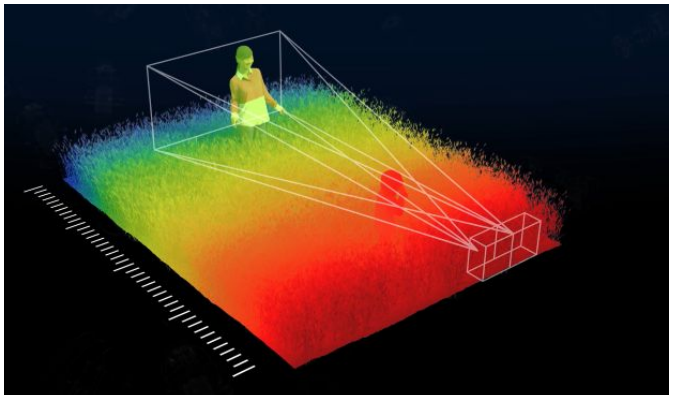
04 Resultados

05 Trabajo Futuro

06 Conclusiones

Introducción

Imágenes de profundidad



[1] Real-Moreno et al. Fast template match algorithm for spatial object detection using a stereo vision system for autonomous navigation. Measurement, 220, 113299.

Aplicaciones

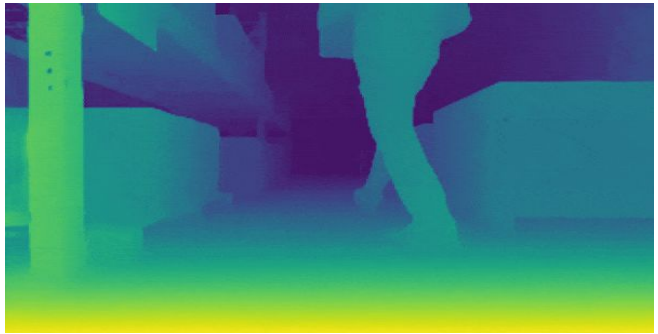
Aplicaciones

Seguridad [2]

[2] Ko, K et al. SqueezeFace: Integrative face recognition methods with LiDAR sensors. Journal of Sensors, 2021(1), 4312245.

Aplicaciones

Seguridad

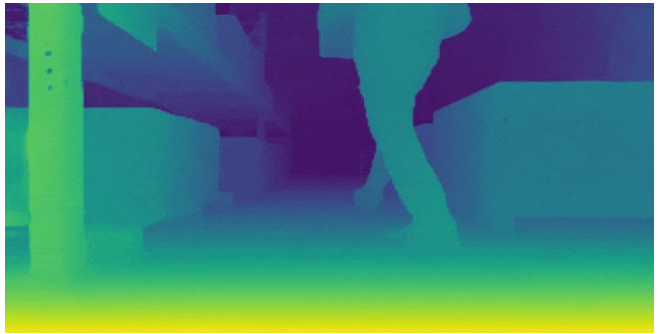


Robótica [3]

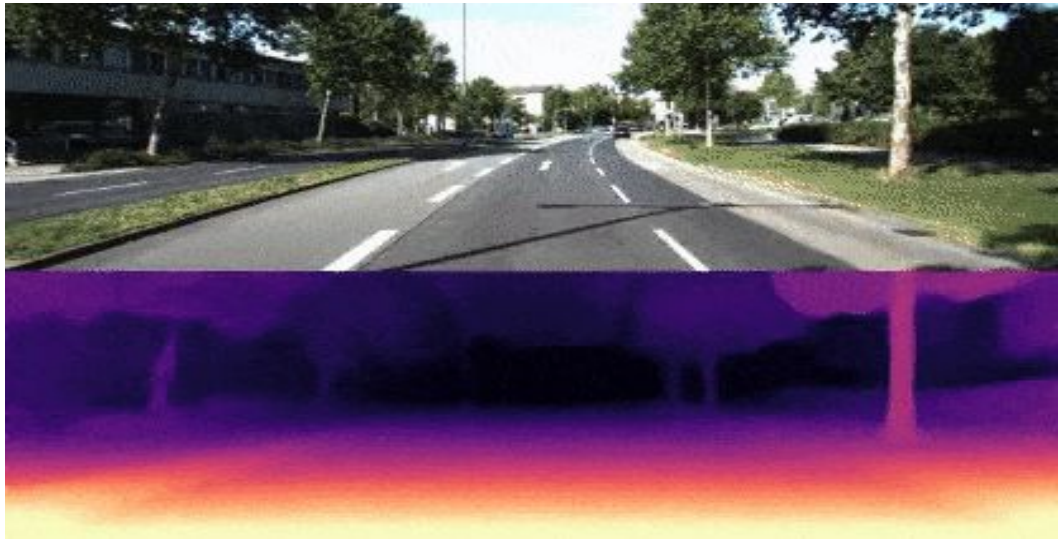
[3] Nalpantidis, L. et al. Stereo vision for robotic applications in the presence of non-ideal lighting conditions. Image and Vision Computing, 28(6), 940-951.

Aplicaciones

Seguridad



Robótica



Vehículos Autónomos [4]

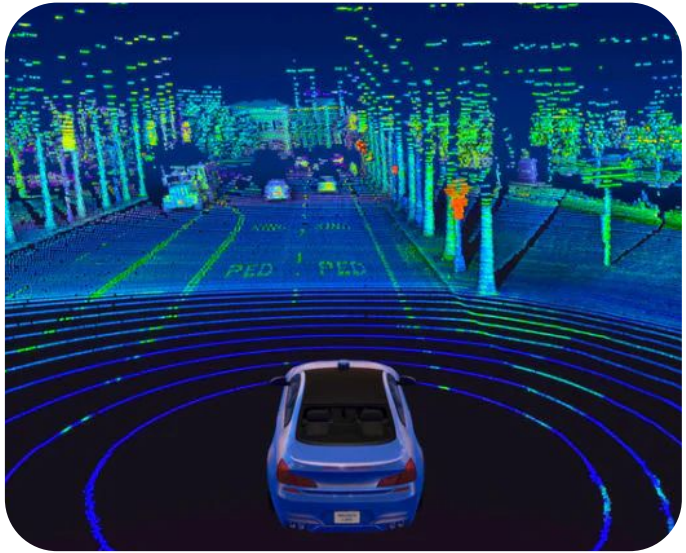
[4] Menze, M. et al. Joint 3d estimation of vehicles and scene flow. ISPRS annals of the photogrammetry, remote sensing and spatial information sciences, 2, 427-434.

Métodos de adquisición

Estereovisión

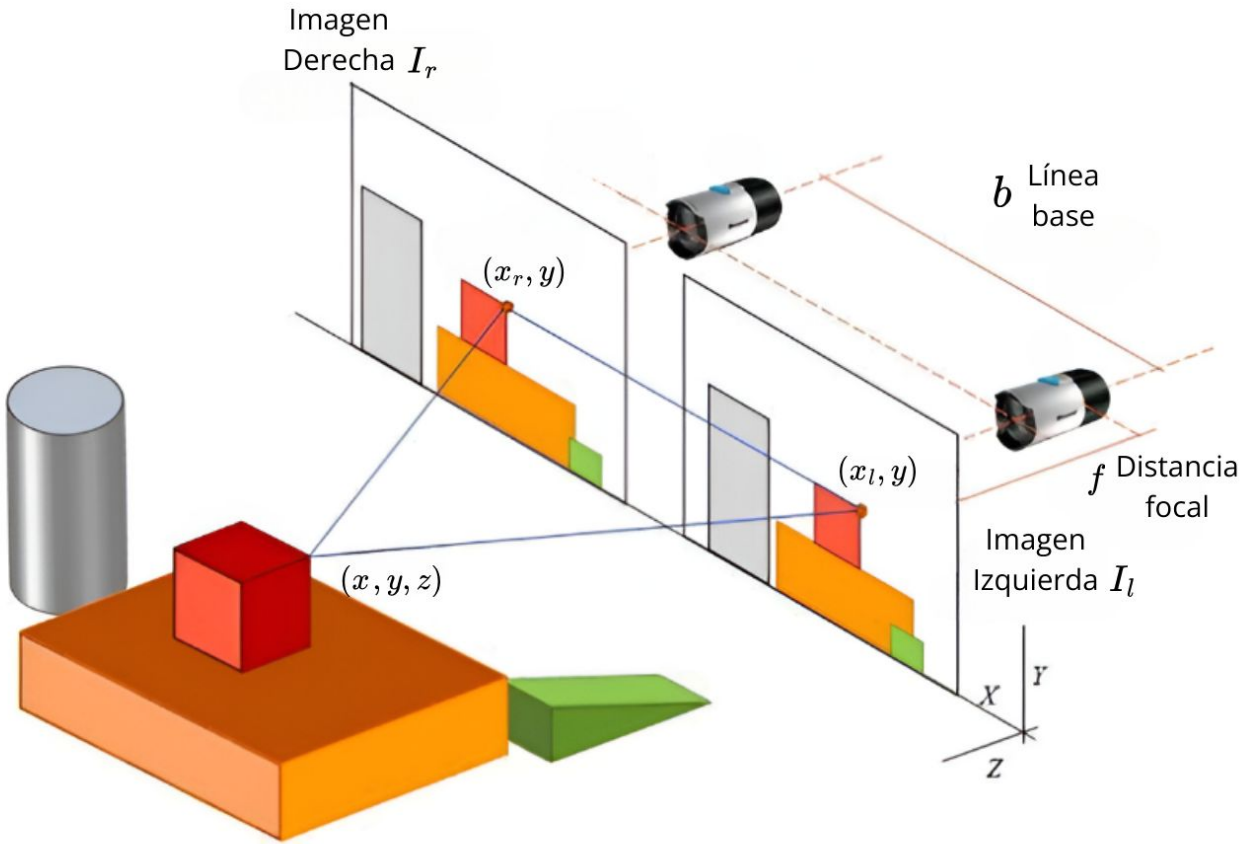
Métodos de adquisición

Estereovisión



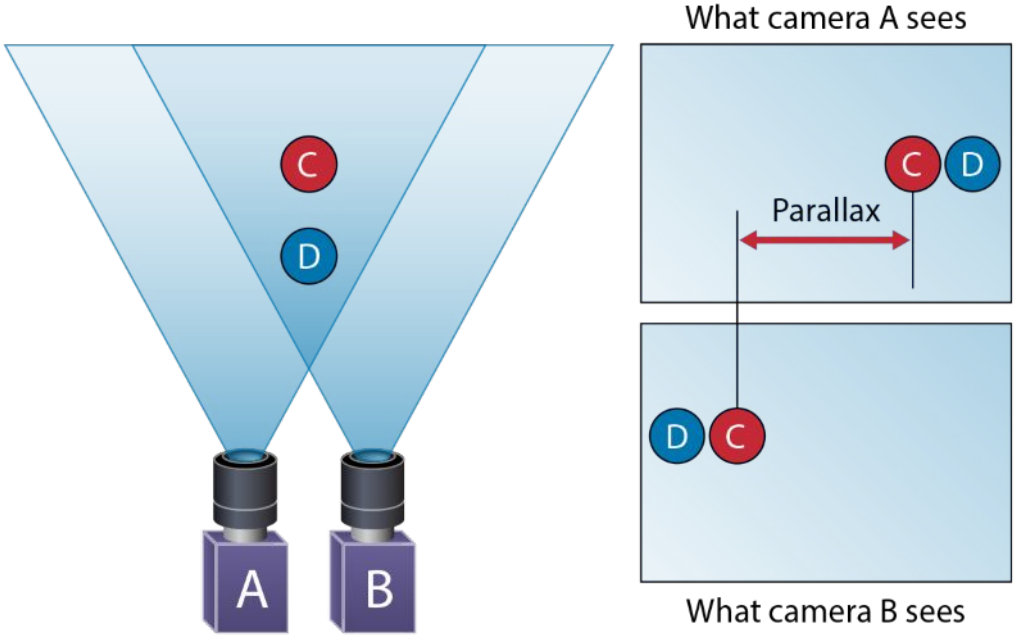
LiDAR

Estereovisión



[5] Colodro-Conde, C. et al. Evaluation of stereo correspondence algorithms and their implementation on FPGA. Journal of Systems Architecture, 60(1), 22-31.

Estereovisión



Disparidad [6]

$$Profundidad = \frac{f \times b}{d}$$

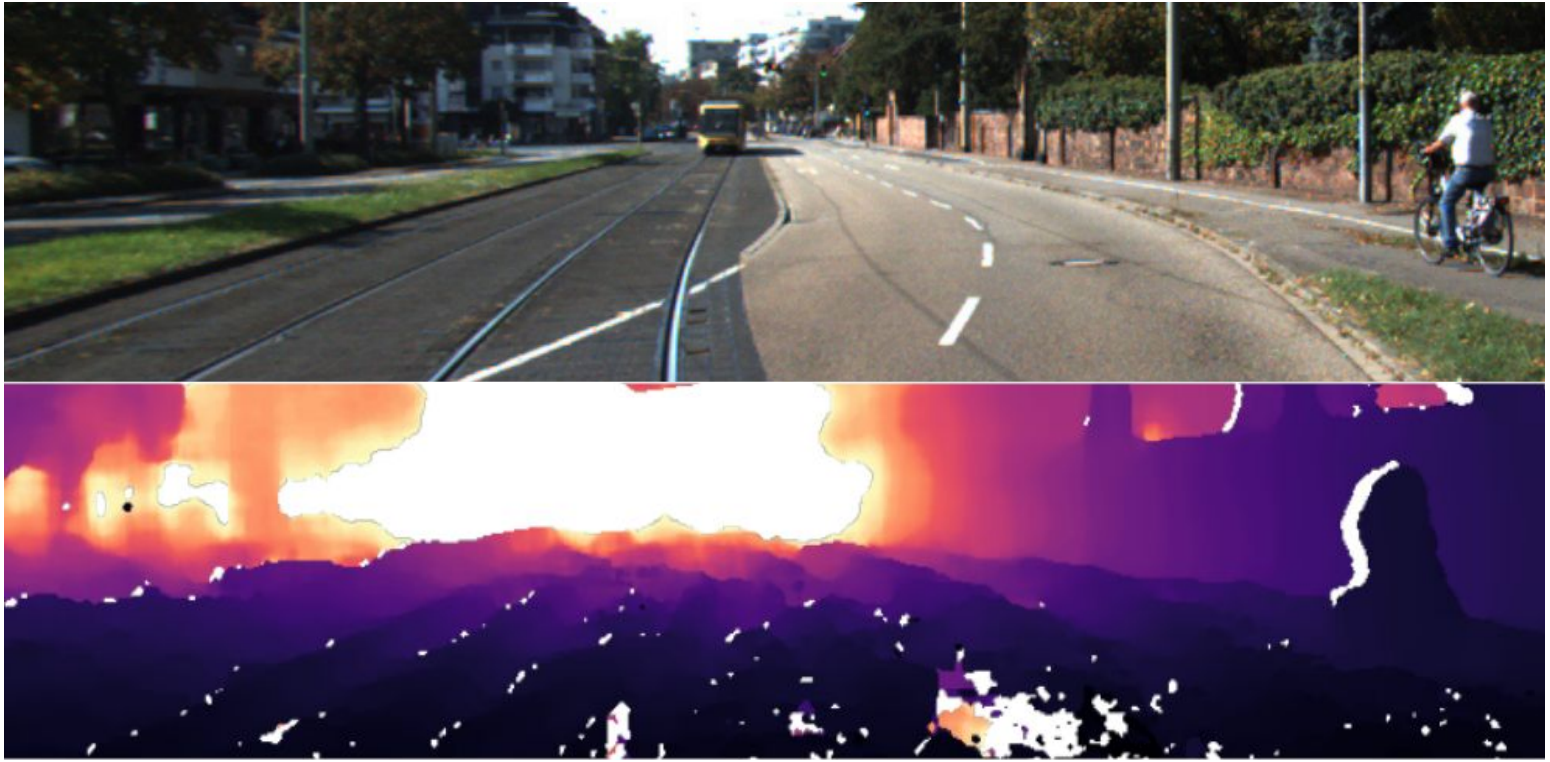
f → Distancia focal

b → Linea base

d → Disparidad

[6] 3-D image processing. Cam2vision. https://www.cam2vision.com/technology-and-capability/comparison-of-speed-metering-technologies/3d-image-processing/

Estereovisión

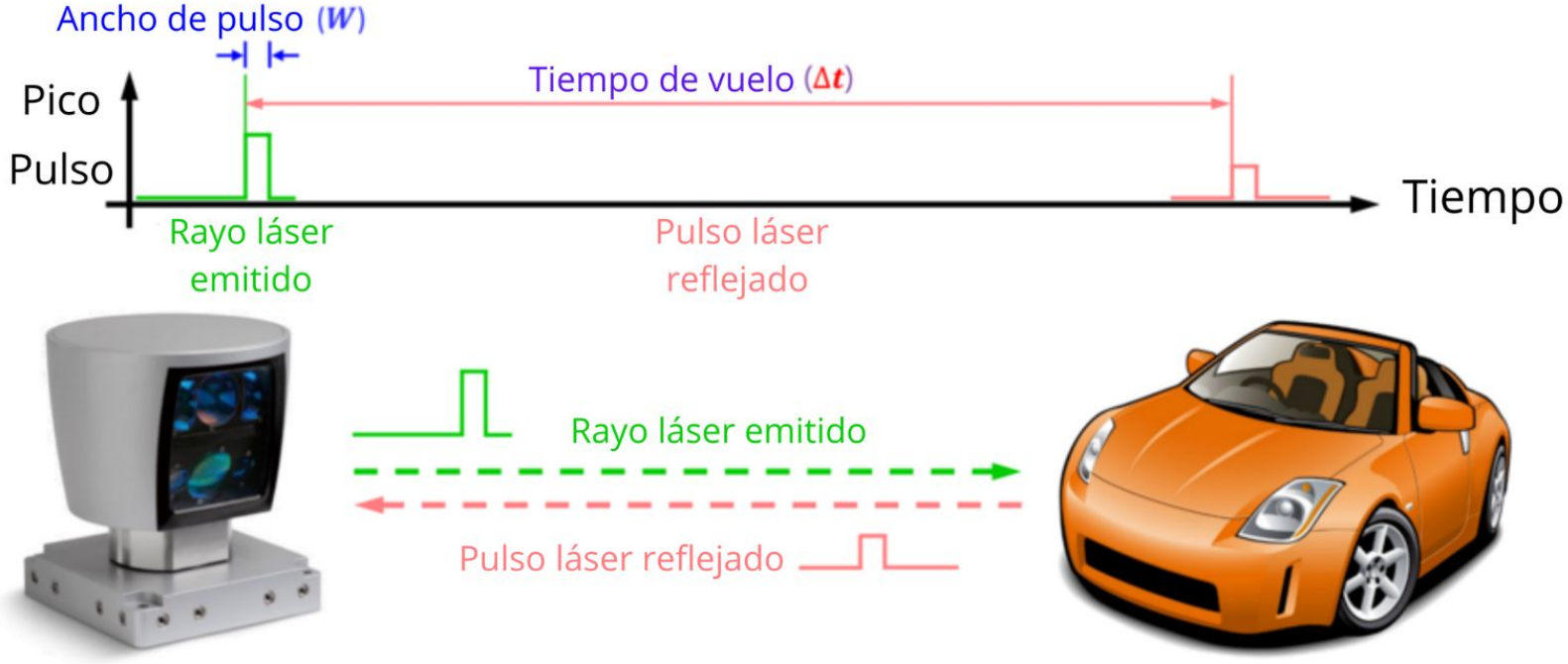


Densidad de puntos

Precisión

Susceptibilidad a errores

LiDAR



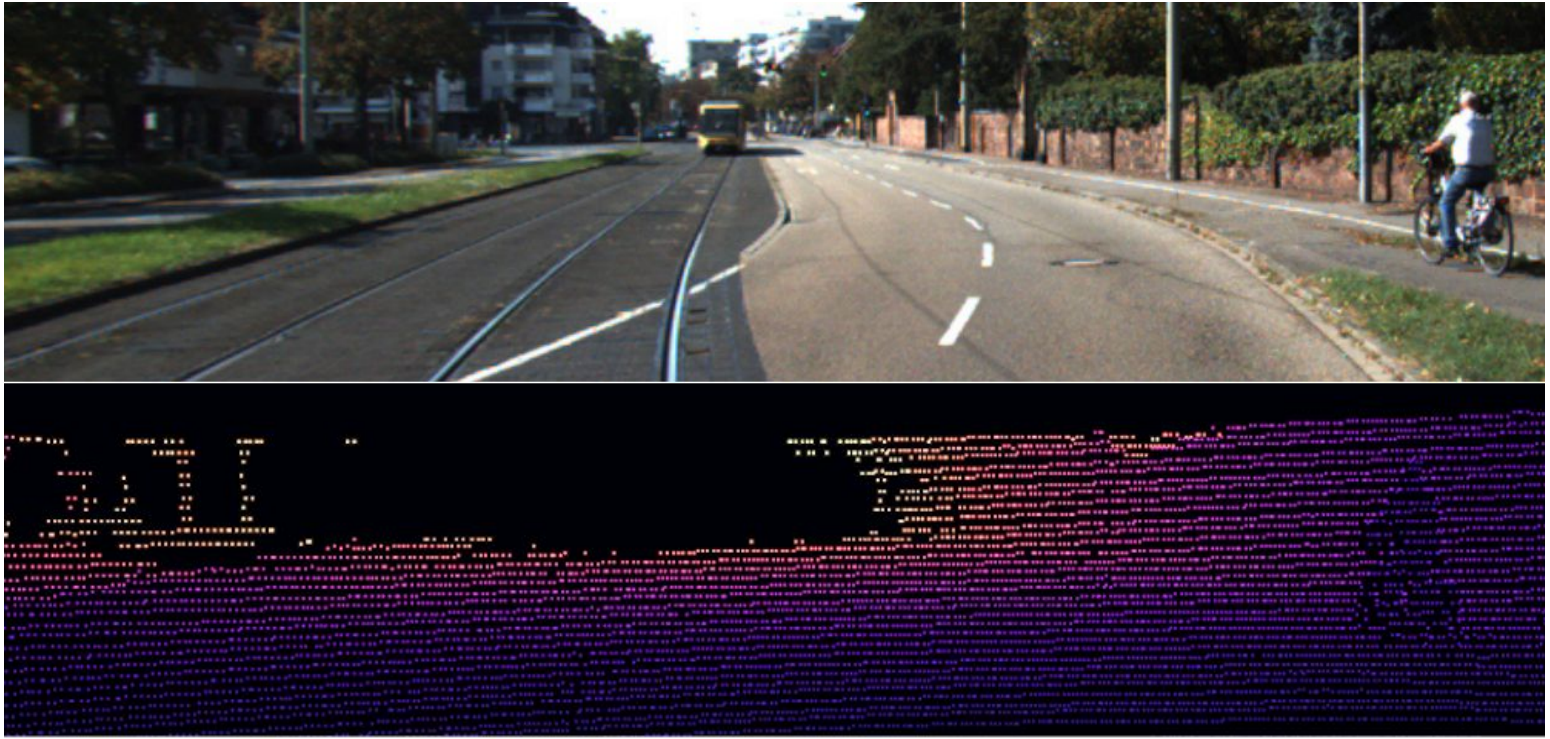
$$Profundidad = \frac{c \times \Delta t}{2}$$

c → Velocidad luz

Δt → Tiempo de vuelo del pulso

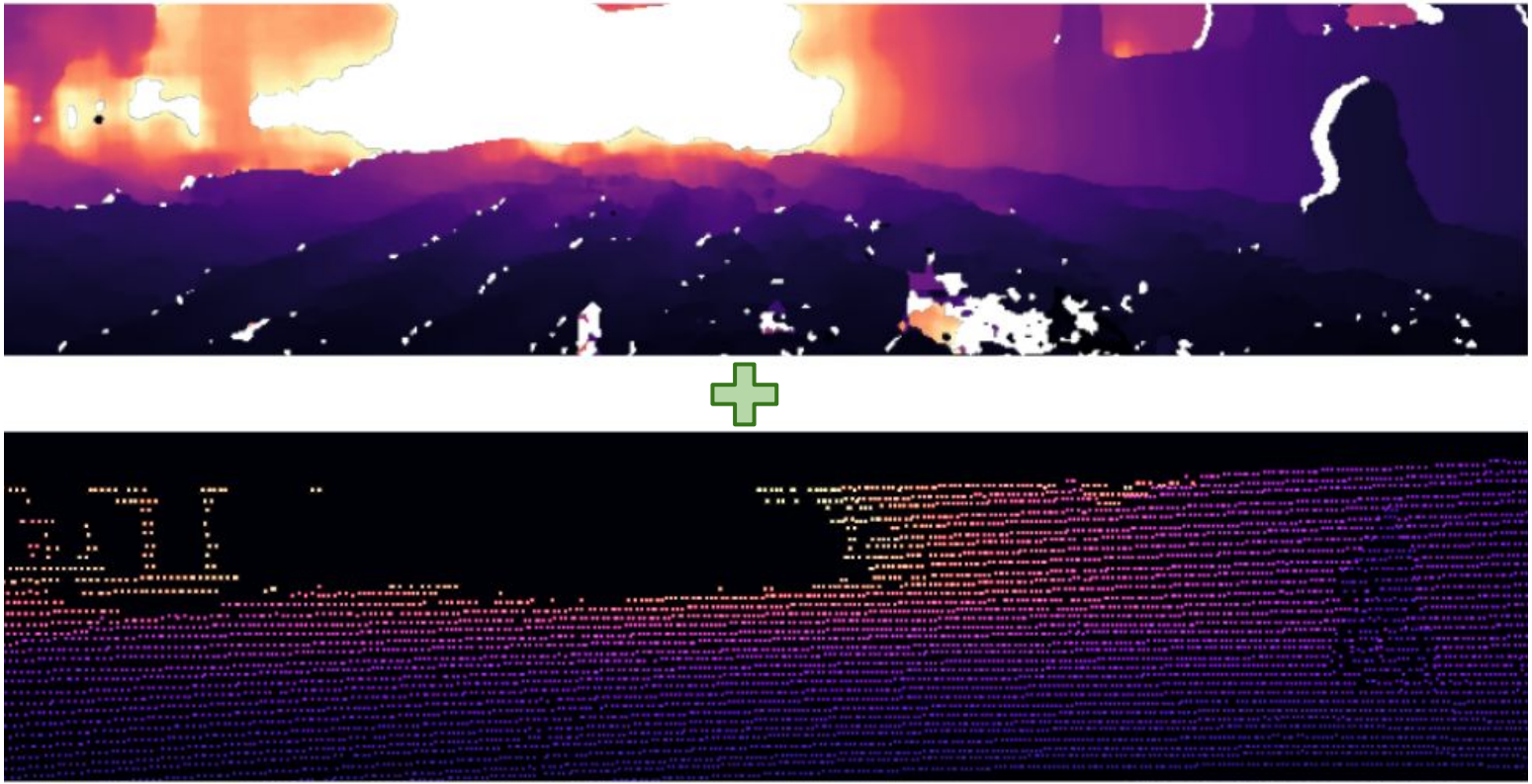
[7] Kim, G. et al. Concurrent firing light detection and ranging system for autonomous vehicles. Remote Sensing, 13(9), 1767.

LiDAR



Precisión **Densidad de puntos** **Costoso**

Lo mejor de ambos mundos



Precisión **Densidad de puntos**

Objetivos

Objetivo General

Desarrollar un algoritmo para la fusión de imágenes de profundidad escasas adquiridas con un sistema de detección y medición de distancia por luz (LiDAR) y densas obtenidas mediante estereovisión, utilizando técnicas de aprendizaje profundo, con el objetivo de mejorar la precisión en la estimación de la profundidad de una escena.

Objetivos Específicos

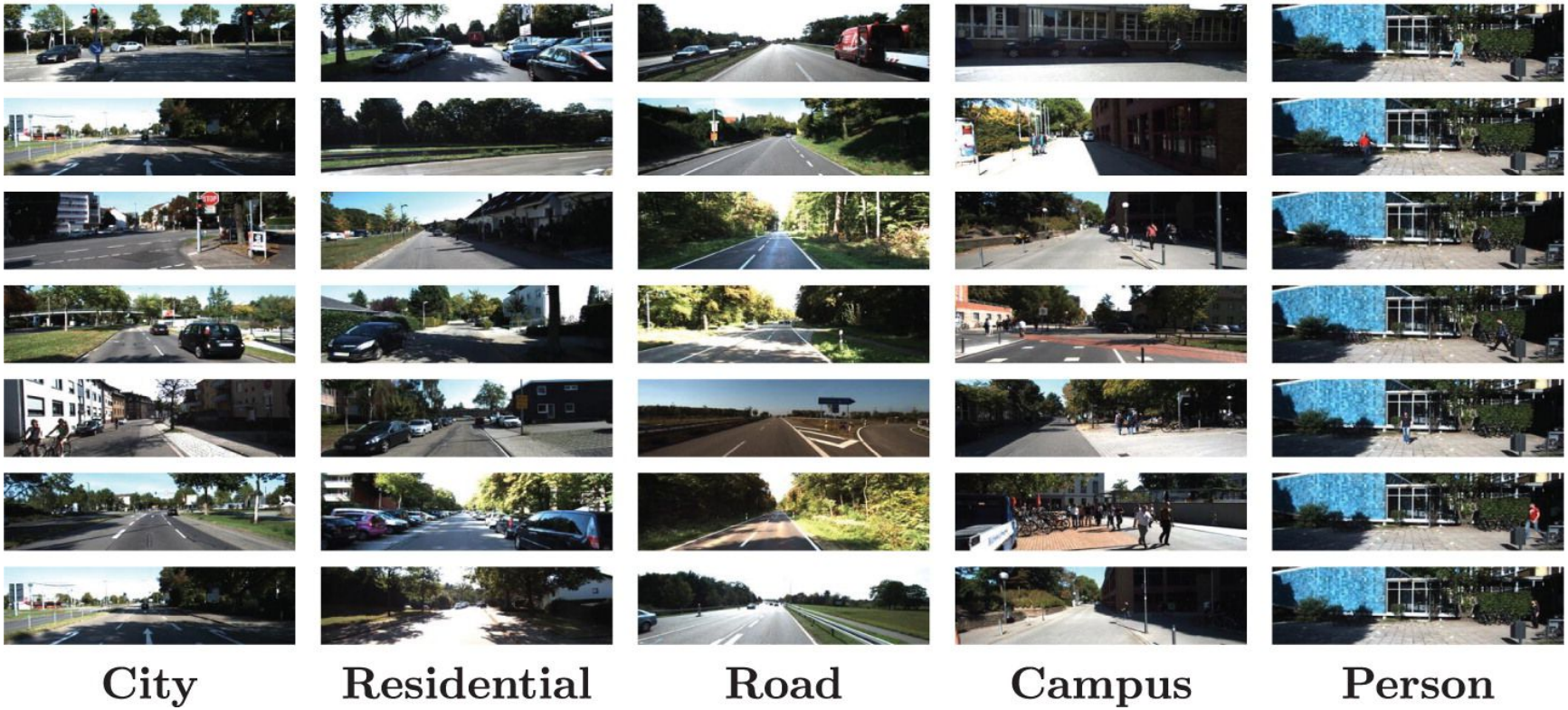
1. Identificar, seleccionar y documentar bases de datos adecuadas que contengan imágenes de profundidad escasas obtenidas con un sistema LiDAR junto con imágenes de profundidad densas adquiridas con un sistema de estereovisión.

2. Diseñar un esquema de fusión de imágenes de profundidad basado en algoritmos de aprendizaje profundo, considerando redes neuronales recurrentes, módulos de atención y transformadores de visión.

3. Implementar en Python un algoritmo computacional para mejorar la precisión de imágenes de profundidad adquiridas con un sistema de estereovisión utilizando imágenes de profundidad de un sistema LiDAR y siguiendo el esquema de fusión propuesto.

4. Evaluar el desempeño del algoritmo desarrollado mediante pruebas con las bases de datos disponibles, comparando los resultados con los algoritmos del estado del arte, específicamente **Stereo-LiDAR Depth Estimation with Deformable Propagation and Learned Disparity-Depth Conversion** y **Holistic and Contextual Evidential Stereo-LiDAR Fusion for Depth Estimation**, en términos de métricas de calidad y precisión de mapas de profundidad.

Dataset KITTI

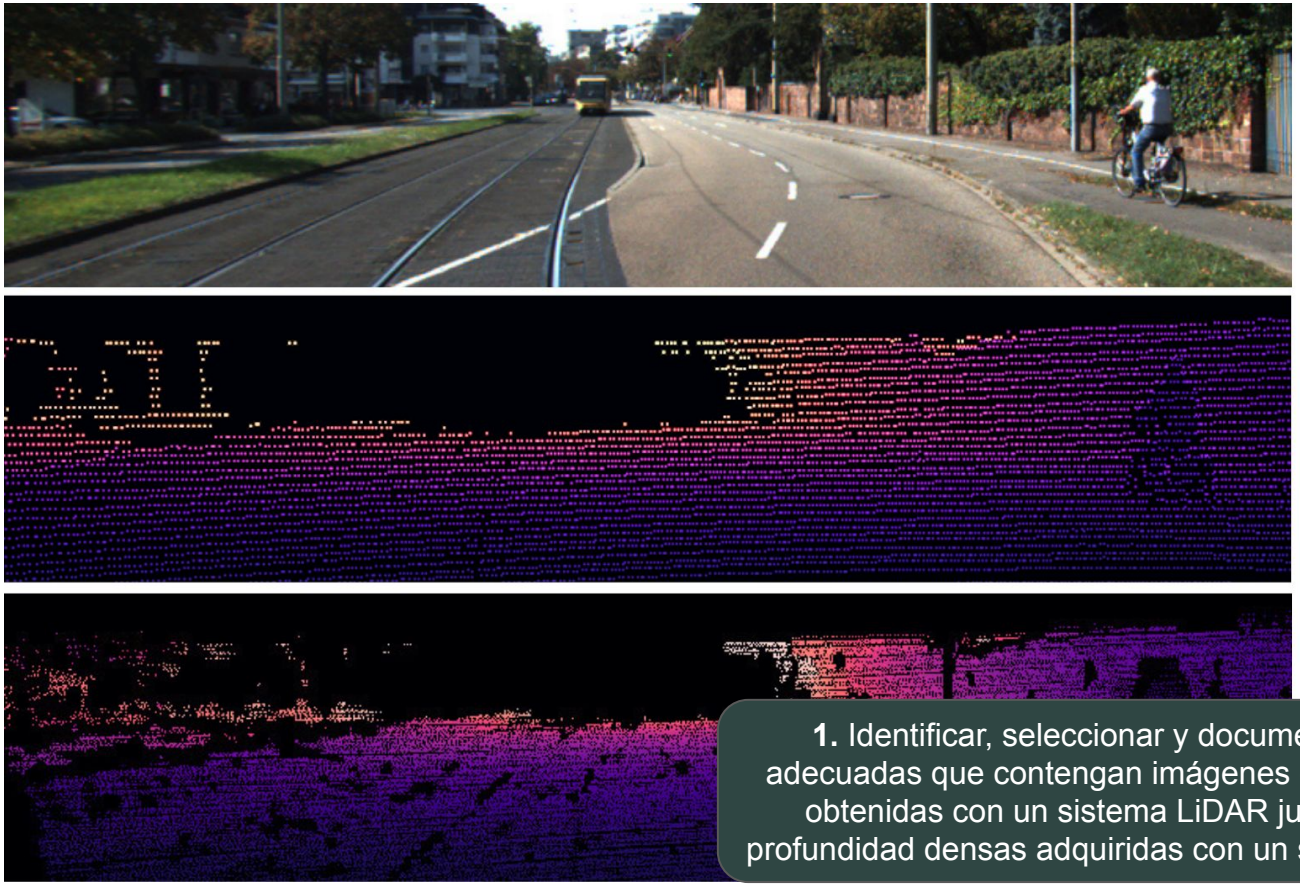


[8] Geiger, A et al. Vision meets robotics: The kitti dataset. The international journal of robotics research, 32(11), 1231-1237.

Dataset KITTI

46.375 Escenas

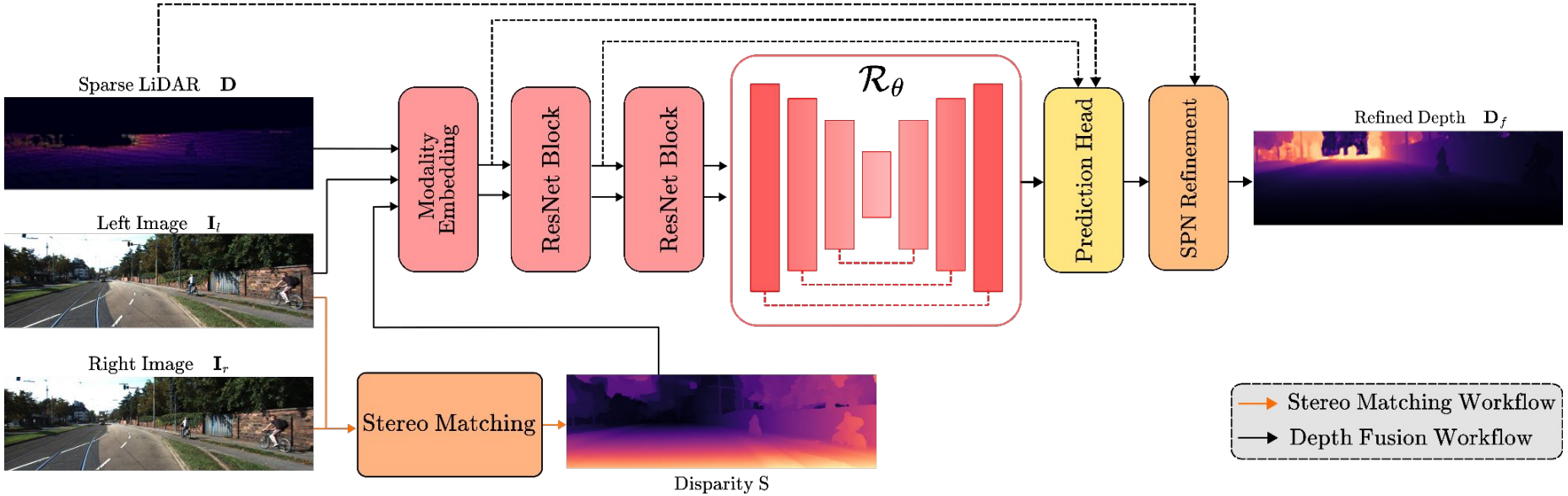
Karlsruher Institut für Technologie



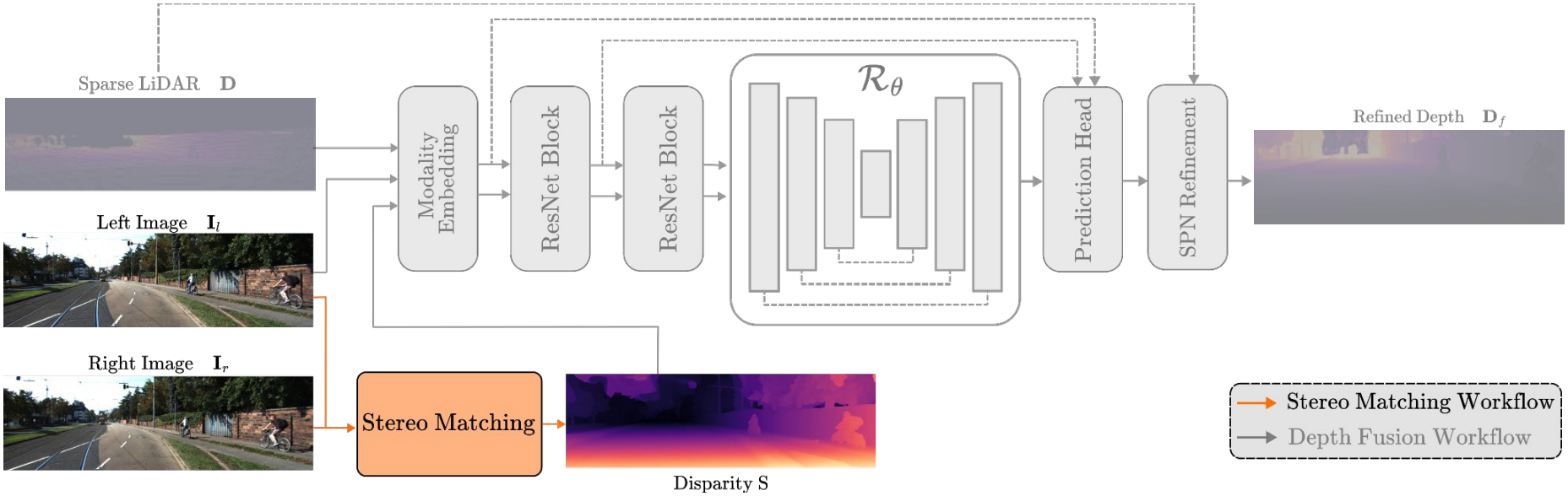
1. Identificar, seleccionar y documentar bases de datos adecuadas que contengan imágenes de profundidad escasas obtenidas con un sistema LiDAR junto con imágenes de profundidad densas adquiridas con un sistema de estereovisión.

Método Propuesto

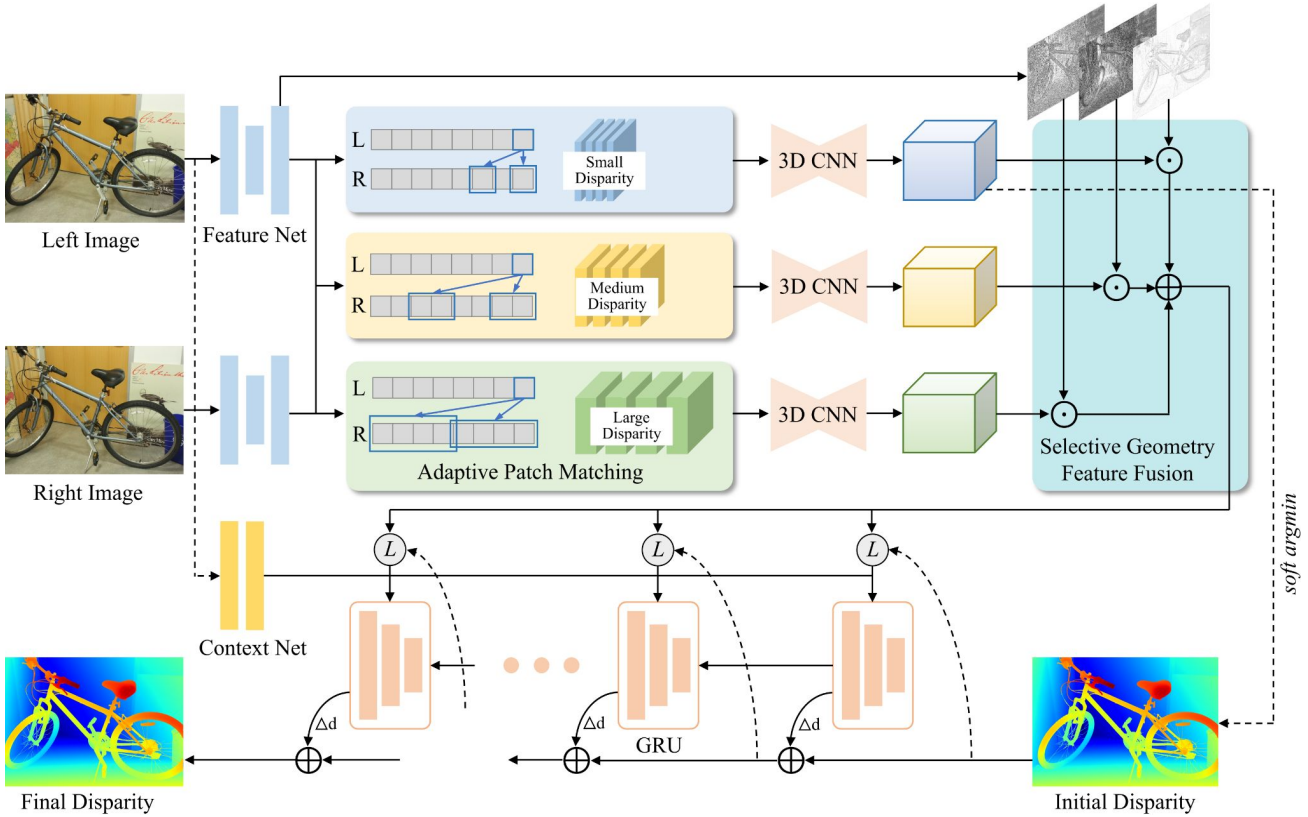
Vista general



Etapa correspondencia estéreo

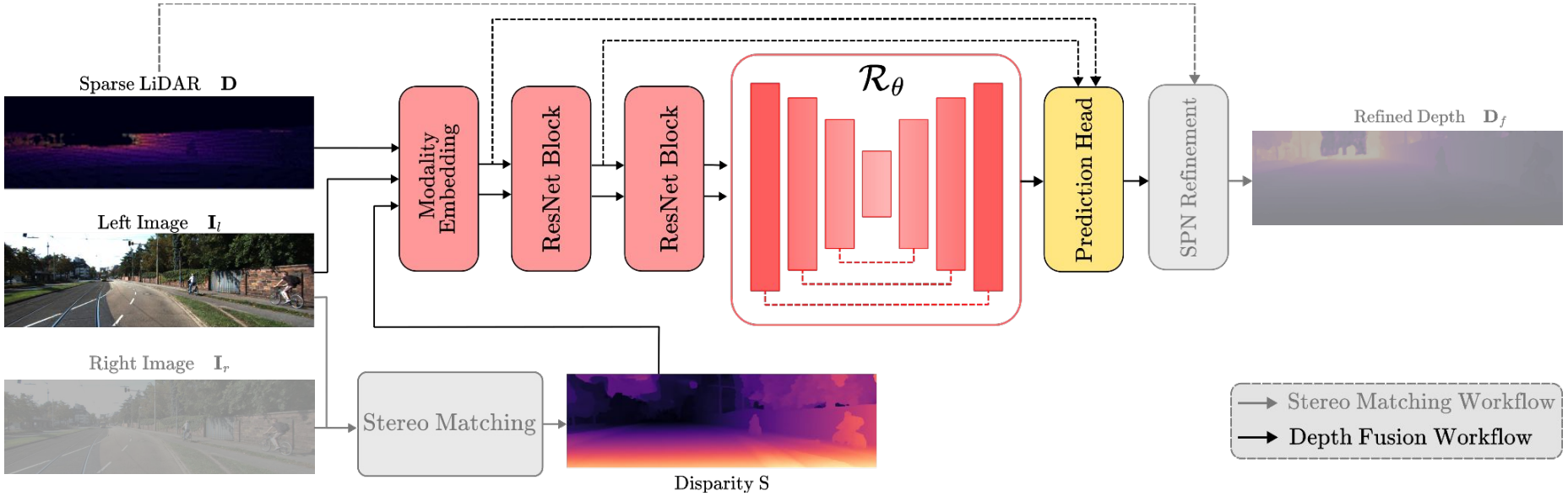


Algoritmo correspondencia estéreo

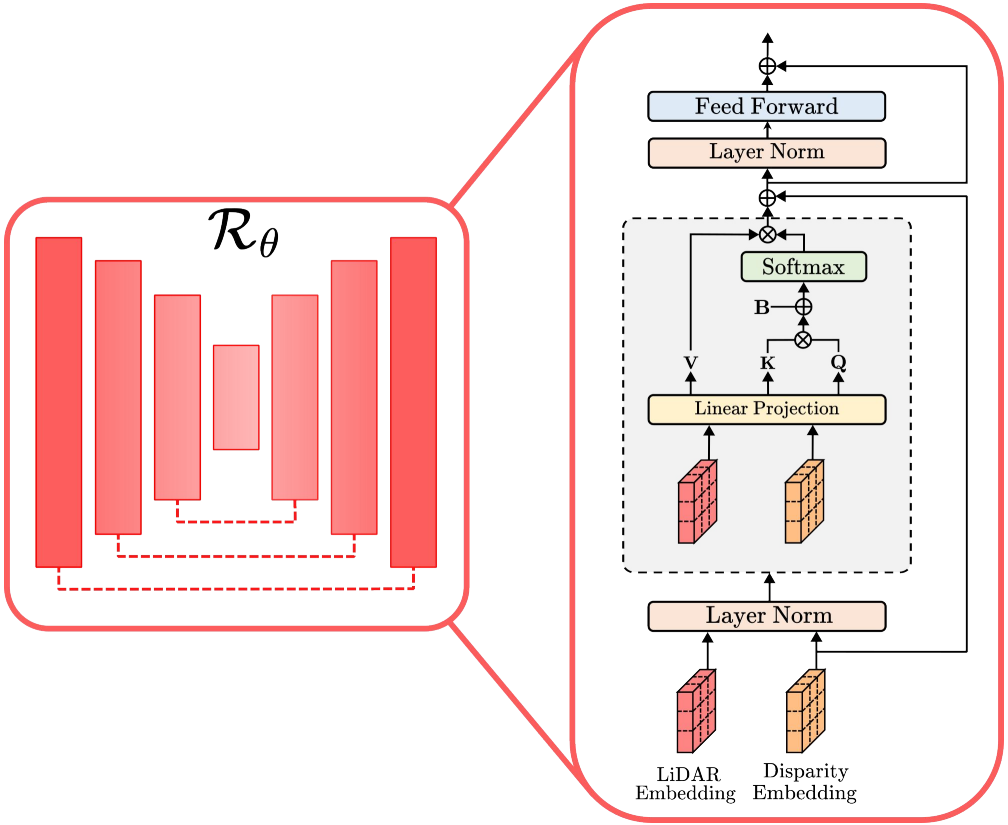


[9] Xu, G. et al. IGEV++: iterative multi-range geometry encoding volumes for stereo matching. arXiv preprint arXiv:2409.00638.

Etapa de Fusión



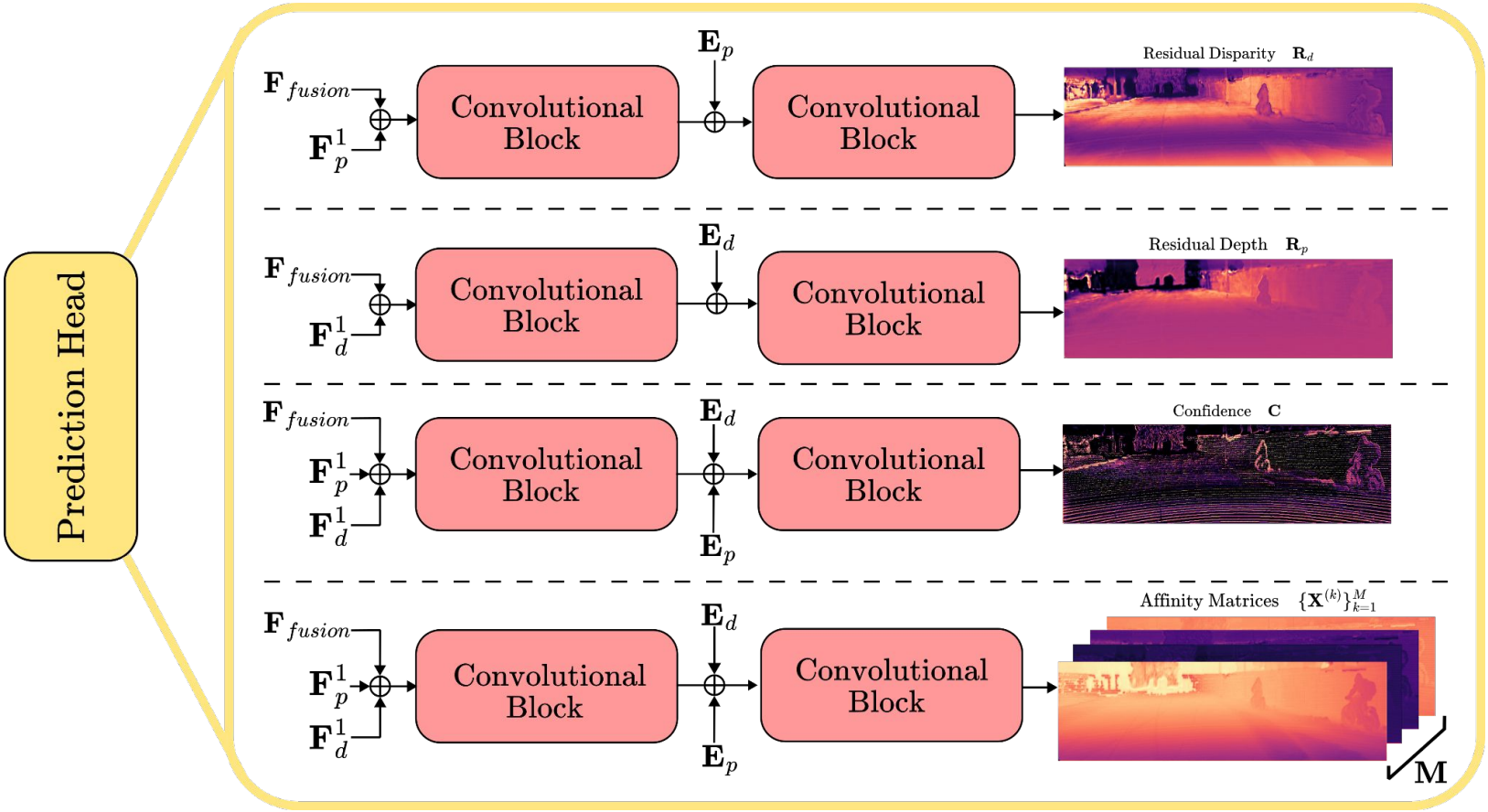
Atención Disparidad-Profundidad



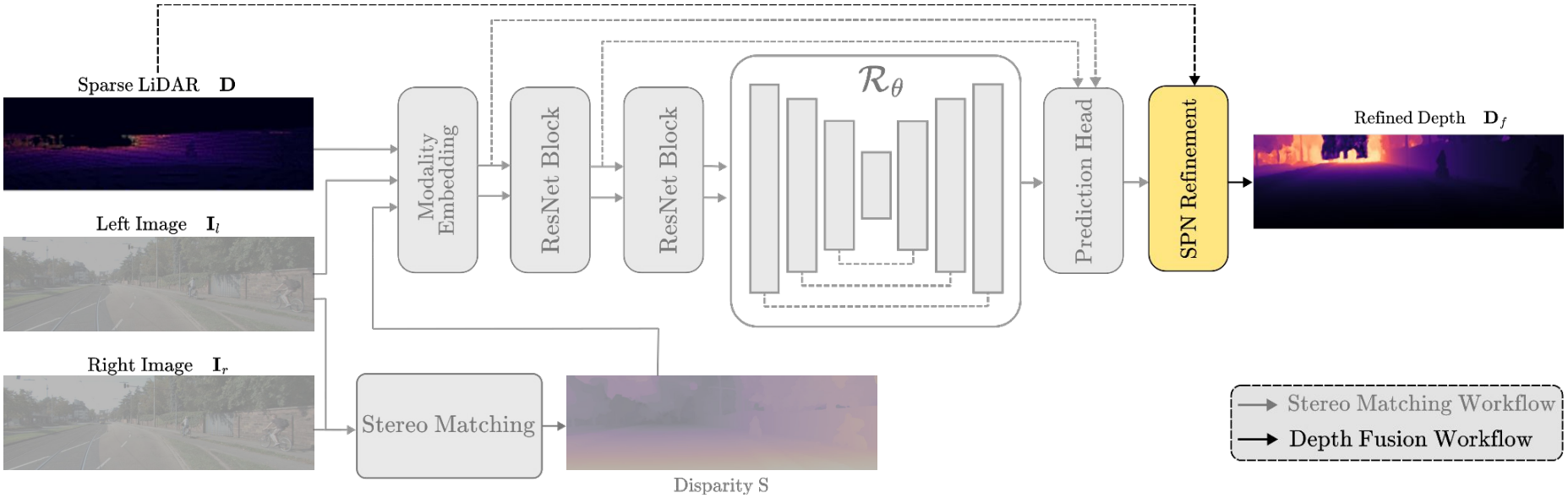
$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

[10] Rho, K. et al. Guideformer: Transformers for image guided depth completion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6250-6259)..

Cabeza de predicción

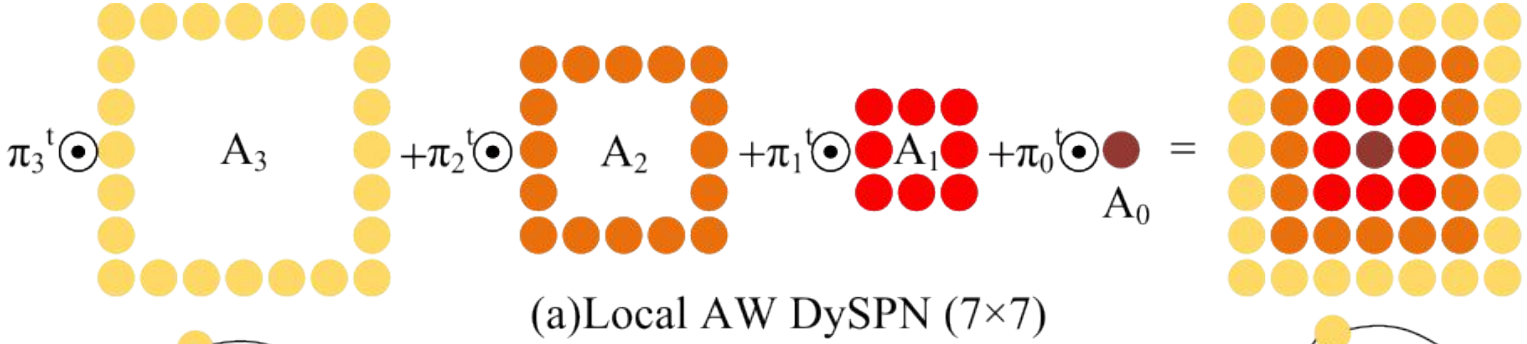


Refinamiento

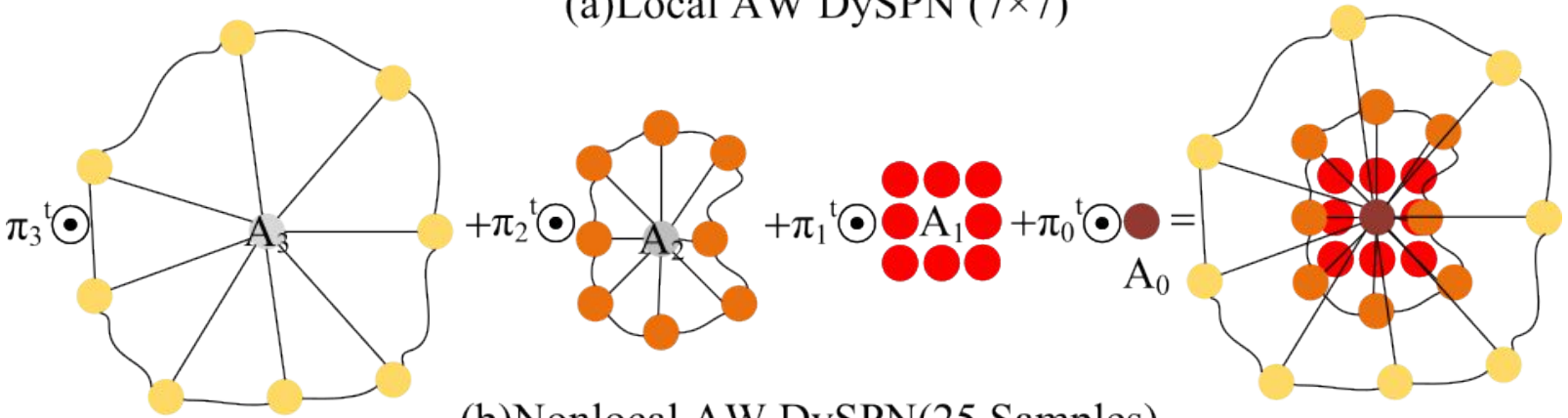


[11] Lin, Y. et al. Dyspn: Learning dynamic affinity for image-guided depth completion. IEEE Transactions on Circuits and Systems for Video Technology, 34(6), 4596-4609.

Refinamiento



(a)Local AW DySPN (7×7)



(b)Nonlocal AW DySPN(25 Samples)

Funciones de pérdida

$$L_{disp}(S, S_{gt}) = \frac{1}{|V|} \sum_{v \in V} \left(|S(v) - S_{gt}(v)| + |S(v) - S_{gt}(v)|^2 \right)$$

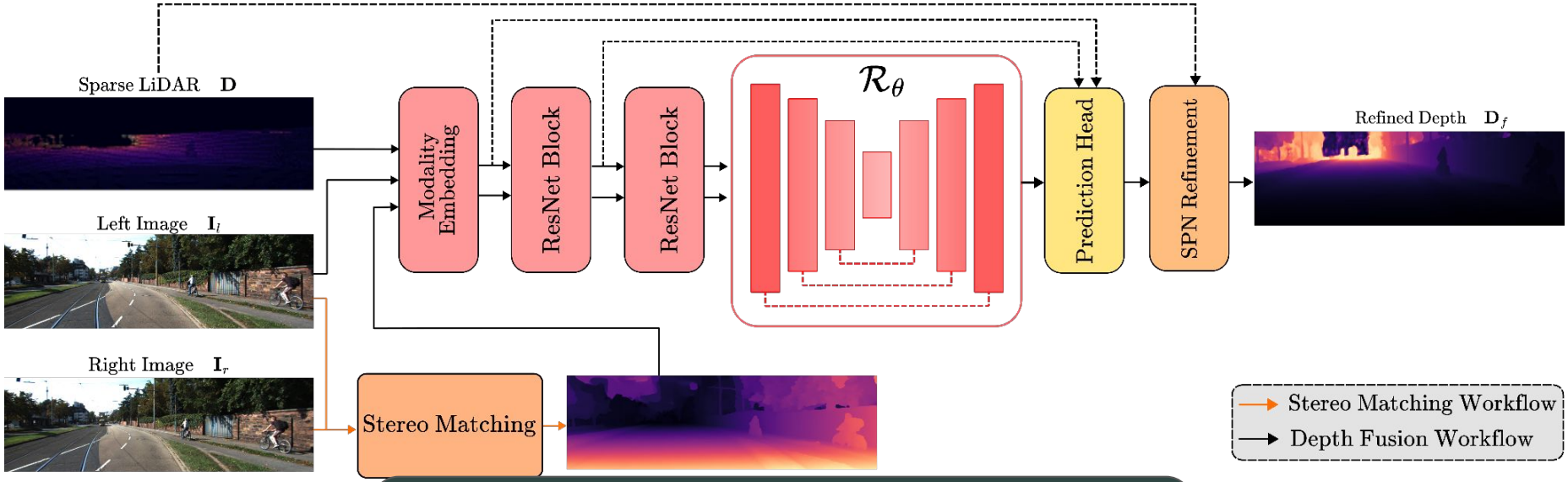
$$L_{fusion}(D_f, D_{gt}) = \frac{1}{|V|} \sum_{v \in V} \left(|D_f(v) - D_{gt}(v)| + |D_f(v) - D_{gt}(v)|^2 \right)$$

$$L_{ref}(\hat{D}_f, D_{gt}) = \frac{1}{|V|} \sum_{v \in V} \left(|\hat{D}_f(v) - D_{gt}(v)| + |\hat{D}_f(v) - D_{gt}(v)|^2 \right)$$

V representa los valores válidos en D_{gt} y S_{gt},
y $|\mathrm{V}|$ denota el tamaño del conjunto de dichos valores.

$$L_{total} = L_{disp} + L_{fusion} + L_{ref}$$

Vista general



2. Diseñar un esquema de fusión de imágenes de profundidad basado en algoritmos de aprendizaje profundo, considerando redes neuronales recurrentes, módulos de atención y transformadores de visión.

Resultados Cuantitativos

Comparación contra SoTA

Método	Entradas	RMSE (mm) ↓	MAE (mm) ↓	iRMSE (1/km) ↓	iMAE (1/km) ↓
Listereo [43]	S+L	832.2	283.91	2.190	1.100
GSM [44]	S+L	793.4	271.48	1.531	0.864
CCVN [45]	S+L	749.3	252.50	**1.397**	0.807
S3 [46]	S+L	703.7	239.60	1.540	0.790
SLFNet [47]	S+L	641.1	197.00	1.773	0.876
VPN [48]	S+L	636.2	205.10	1.872	0.987
EG-Depth [49]	S+L	675.5	197.16	1.600	0.787
SDG-Depth[8]	S+L	<u>623.2</u>	197.55	1.519	**0.772**
HCENet[12]	S+L	**599.3**			
Nuestro	S+L	625.4			

Cuadro 1. Comparación de rendimiento
senta estereovisión junto con LiDAR, respectivamente.

3. Implementar en Python un algoritmo computacional para mejorar la precisión de imágenes de profundidad adquiridas con un sistema de estereovisión utilizando imágenes de profundidad de un sistema LiDAR y siguiendo el esquema de fusión propuesto.

A diferentes rangos de profundidad

Rango de profundidad	RMSE (mm) ↓	MAE (mm) ↓	iRMSE (1/km) ↓	iMAE (1/km) ↓
0-20 m	227.2	94.50	1.688	0.871
20-50 m	973.8	408.13	1.170	0.463
50-100 m	2594.7	1148.63	1.030	0.347

Cuadro 3. Comparación de métricas de rendimiento en diferentes rangos de profundidad.

Ablaciones

Configuración del modelo			Métricas de rendimiento				# Parámetros (M)
Correspondencia stereo	Fusion ViT	Refinamiento	RMSE (mm) ↓	MAE (mm) ↓	iRMSE (1/km) ↓	iMAE (1/km) ↓	
✓	✗	✗	915.9	340.42	1.740	1.048	14.52
✓	✓	✗	647.6	202.63	1.618	0.858	99.74
✓	✓	✓	625.4	180.88	1.604	0.773	99.74

Cuadro 4. Comparación del uso de módulos, métricas de rendimiento y número de parámetros en distintas configuraciones de modelo.

Comparación porcentual

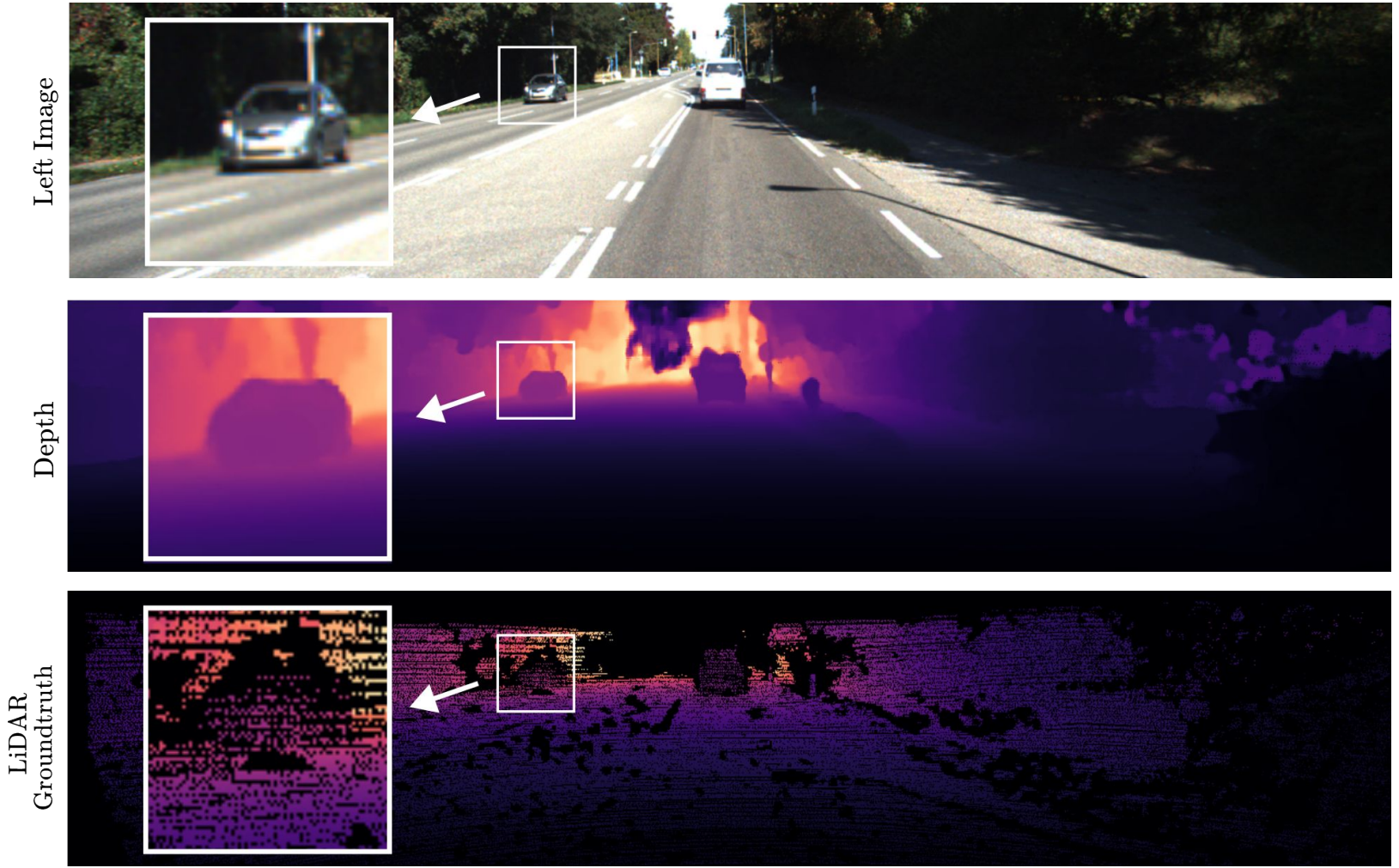
Método	RMSE (mm) ↓	MAE (mm) ↓	iRMSE (1/km) ↓	iMAE (1/km) ↓
SDG-Depth[8]	0.35 %	8.43 %	5.59 %	0.13 %
HCENet[12]	4.35 %	4.80 %	11.45 %	0.90 %

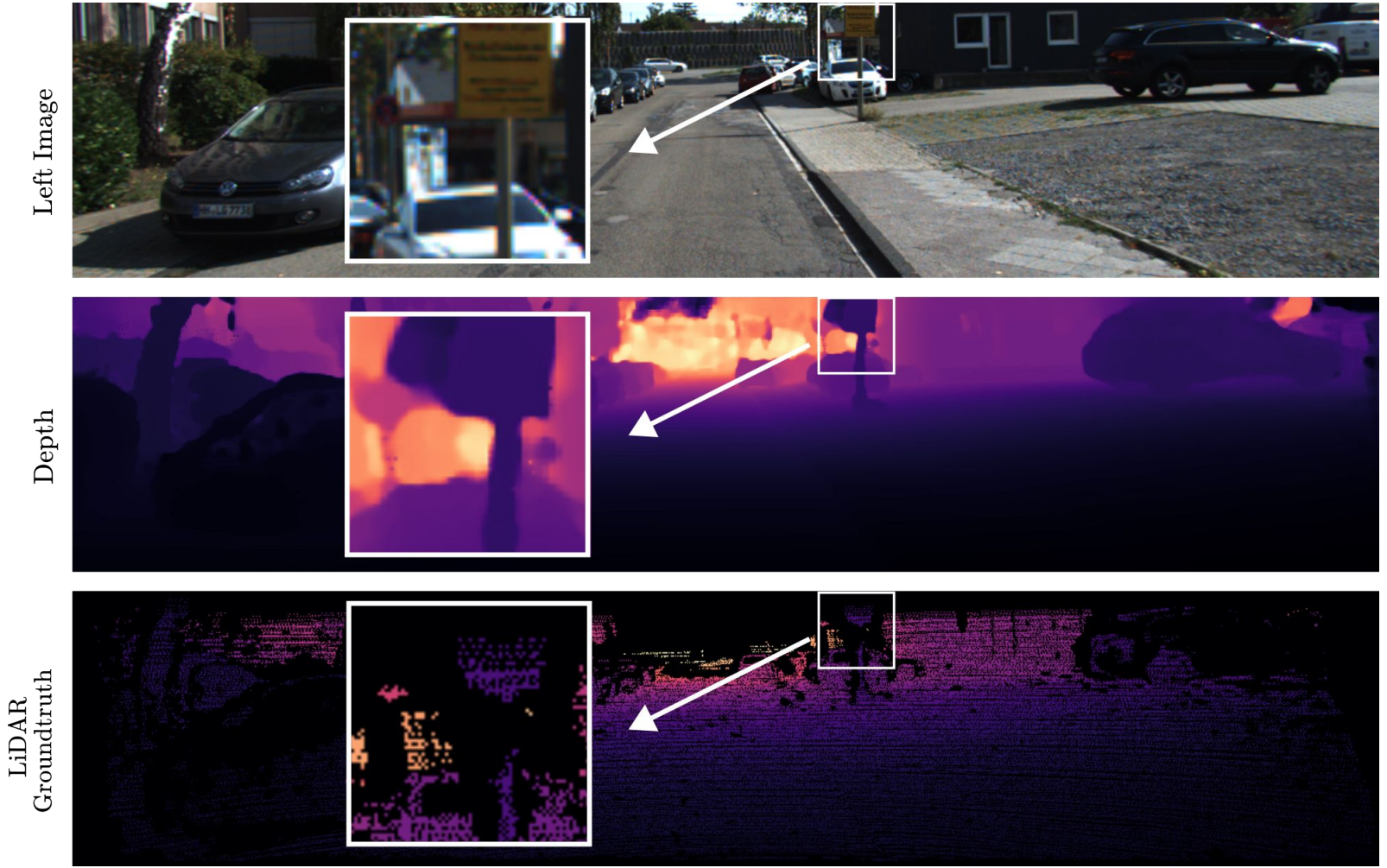
4. Evaluar el desempeño del algoritmo desarrollado mediante pruebas con las bases de datos disponibles, comparando los resultados con los algoritmos del estado del arte, específicamente SDG-Depth y HCENet, en términos de métricas de calidad y precisión de mapas de profundidad.

[12] Li, A. et al. Stereo-lidar depth estimation with deformable propagation and learned disparity-depth conversion. In 2024 IEEE ICRA (pp. 2729-2736). IEEE.

[13] Fan, J. et al. Holistic and contextual evidential stereo-LiDAR fusion for depth estimation. IEEE Transactions on Intelligent Vehicles.

Resultados Cualitativos


Left Image
Depth
LiDAR
Groundtruth


Left Image
Depth
LiDAR
Groundtruth

Vista tridimensional

Figura 15. Visualización de una imagen RGB convertida en una representación tridimensional mediante una nube de puntos, a partir de un mapa de profundidad generado por nuestro método.

Conclusiones

Conclusiones

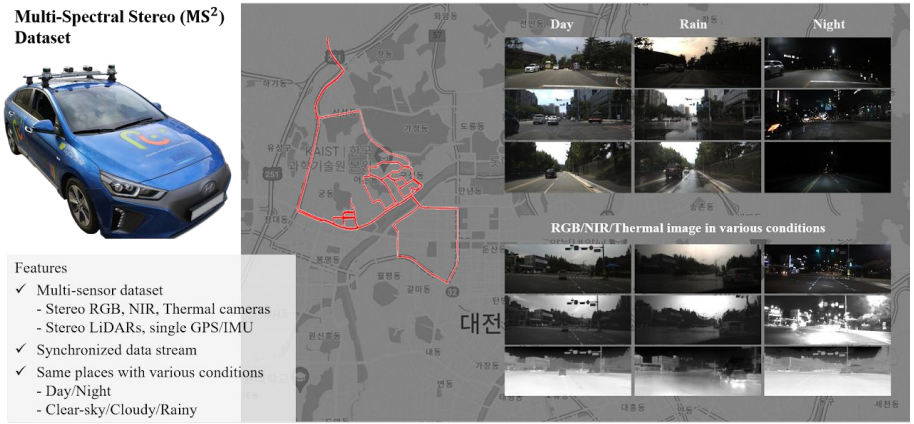
El método propuesto aprovecha las ventajas de ambas fuentes para obtener imágenes de profundidad densas y precisas.

Los experimentos realizados evidencian la competitividad de nuestro algoritmo frente a los métodos del estado del arte, destacando un MAE de *180.88 mm*, que representa el valor más alto entre los evaluados

Demostramos la eficacia de utilizar algoritmos pre-entrenados en correspondencia estéreo para completar datos LiDAR, permitiendo que una arquitectura basada en la atención establezca correspondencias disparidad-profundidad que mejoran la propagación y la estimación de la profundidad.

Trabajo Futuro

Trabajo Futuro



MS2 Dataset [14]

Modelos Fundacionales

[14] Shin, U. et al. Deep depth estimation from thermal image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1043-1053).

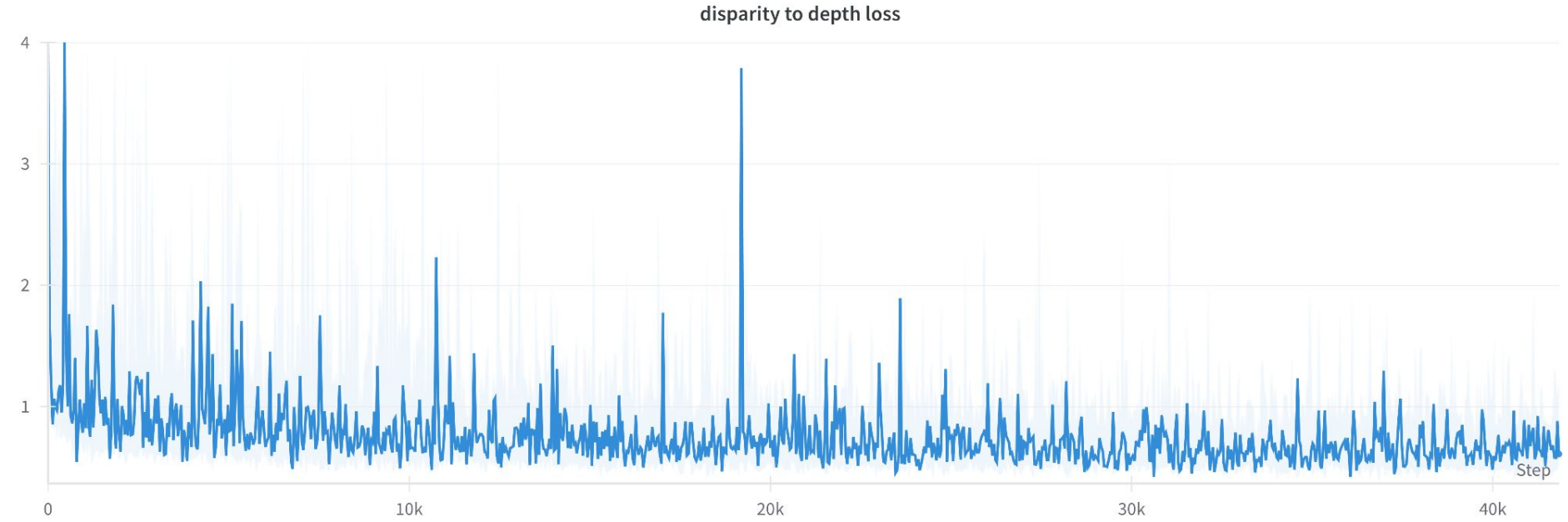
CoICACI 2025

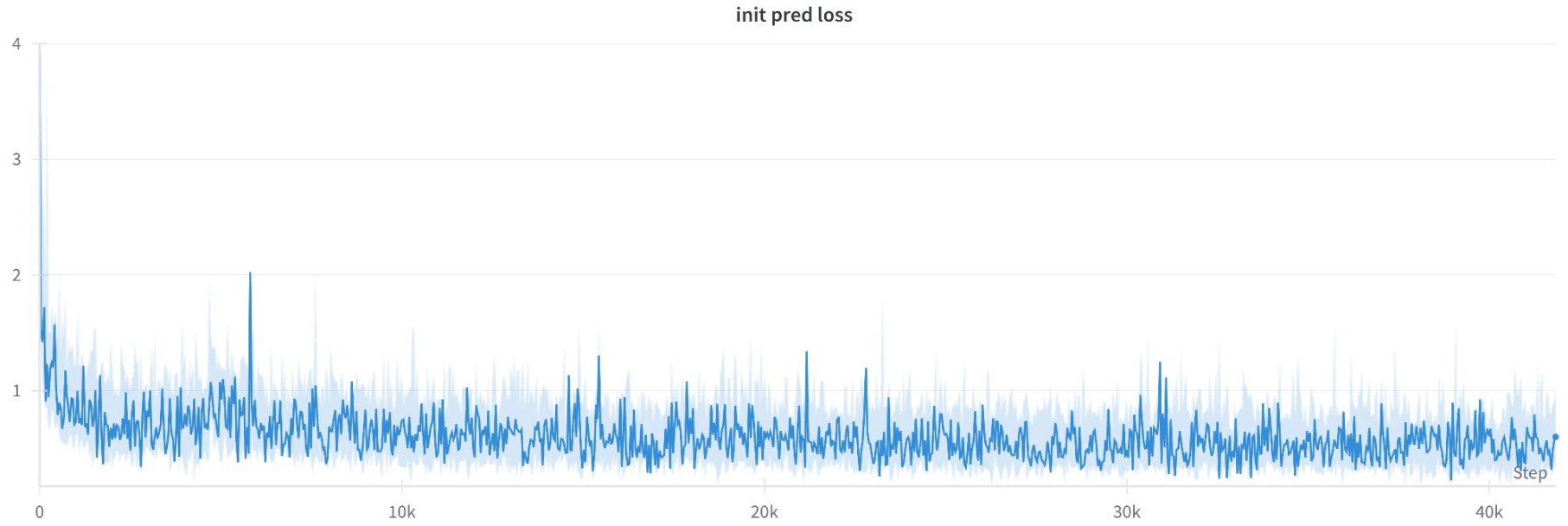
COLOMBIAN CONFERENCE ON APPLICATIONS OF COMPUTATIONAL INTELLIGENCE

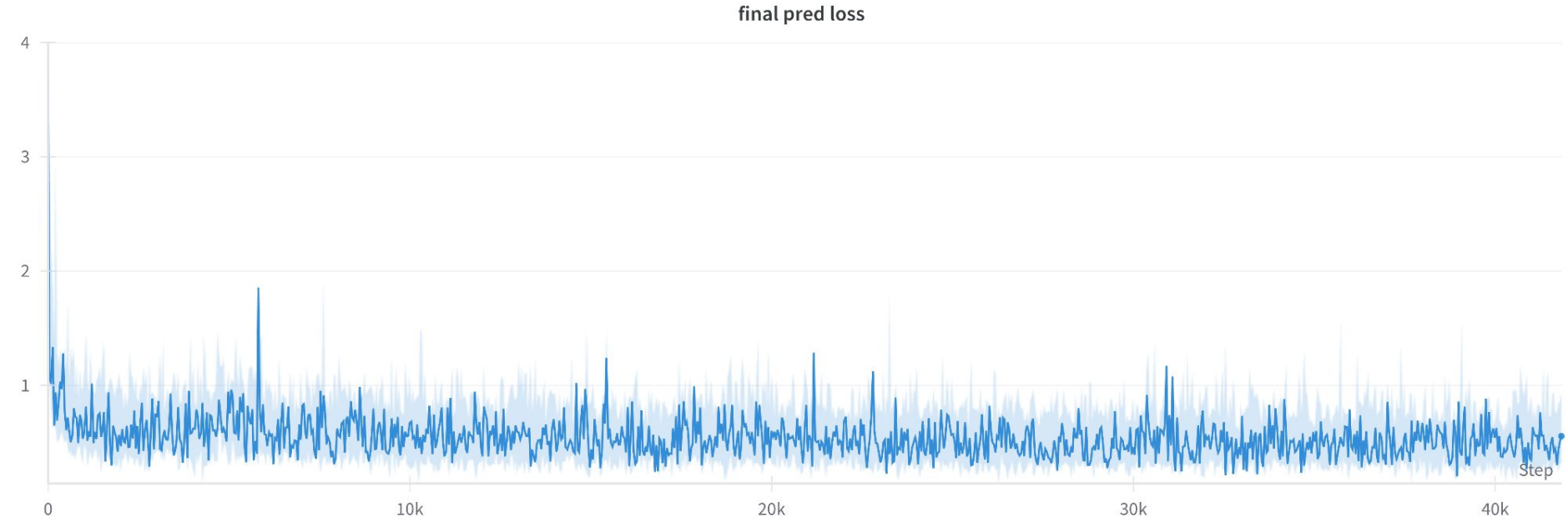
August 27th - 29th, 2025 - Armenia, Colombia

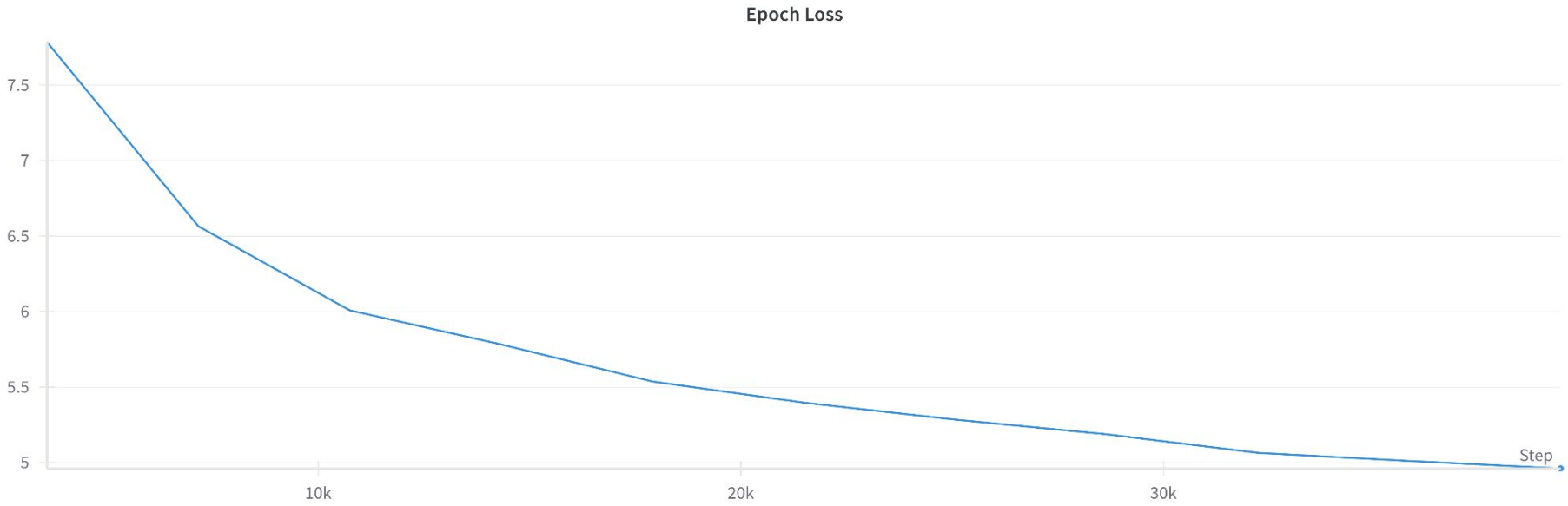
Gracias

Anexos


disparity to depth loss
4
3
2
1
0
10k
20k
30k
40k
Step


init pred loss
4
3
2
1
0
10k
20k
30k
40k
Step


final pred loss
4
3
2
1
0
10k
20k
30k
40k
Step


Epoch Loss
7.5
7
6.5
6
5.5
5
Step
10k
20k
30k