# S²M²: Scalable Stereo Matching Model for Reliable Depth Estimation

Junhong Min[1]*, Youngpil Jeon[1], Jimin Kim[1], Minyong Choi[1]
[1]Samsung Electronics, *Corresponding author

ICCV OCT 19-23, 2025 HONOLULU HAWAII

## Motivation & Objective

Prior stereo matching models struggled to generalize across diverse input conditions. Attempts to scale models often led to inefficiencies, revealing a need for a more adaptable solution.

**We aim to develop a unified architecture that achieves**

- **Input Scalability**: Robust performance across varying image resolutions and disparity ranges.
- **Model Scalability**: Consistent performance gains with increased model capacity.

## Our Key Contributions

- **Scalable Global Matching Architecture**
  A multi-resolution Transformer design that scales effectively with both input complexity and model size, and is fully trainable from scratch.
- **Accurate & Reliable Depth Estimation**
  A novel PMC loss function enhances disparity precision, while joint confidence and occlusion estimation ensures reliable depth maps.
- **New State-of-the-Art Performance**
  Our method, S²M², achieves **top-ranked** performance on challenging real-world benchmarks, including Middlebury v3 and ETH3D.

## Proposed Method:

### Architecture

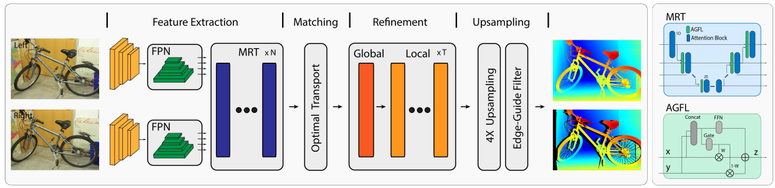
Our proposed model revitalizes the global matching paradigm.

**Powerful Multi-scale Feature Extraction**

- **Multi-Resolution Transformer (MRT)**: Employs a hybrid attention strategy (horizontal at high-res, 2D at low-res) to balance performance and computational cost.
- **Adaptive Gated Fusion Layer (AGFL)**: Acts as a dynamic gate that selectively fuses features across different scales, ensuring robust information exchange.

**Robust Global Matching**

Utilizes Optimal Transport to find a globally consistent matching plan from all possible correspondences.



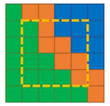

### Loss Function

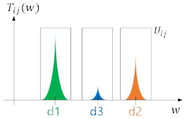
Our model is trained with a composite loss function that combines standard L1 losses with our novel **Probabilistic Mode Concentration (PMC)** loss.

- Since global matching is performed on 1/4-downsampled features, a more direct mechanism is required to guide the matching probabilities.
- PMC loss directly regularizes the matching probability distribution, encouraging it to concentrate on valid disparity candidates.

$$S_{ij} = \sum_{w \in \mathcal{U}_{ij}} T_{ij}(w) \quad \mathcal{U}_{ij} = \bigcup_{k=1}^{K} \mathcal{N}(c_{ij}^{(k)})$$

$$\mathcal{L}_{\text{PMC}} = \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} \max(1 - S_{ij} - \epsilon, 0)$$






The final loss function integrates disparity, confidence, occlusion, and PMC losses, each weighted by a dedicated hyperparameter.

$$\mathcal{L} = \lambda_D \mathcal{L}_D + \lambda_{\mathcal{O}} \mathcal{L}_{\mathcal{O}} + \lambda_{\mathcal{C}} \mathcal{L}_{\mathcal{C}} + \lambda_{\text{PMC}} \mathcal{L}_{\text{PMC}}$$

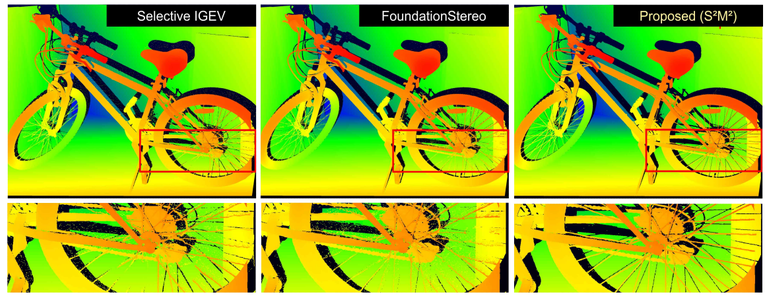
## Experiments

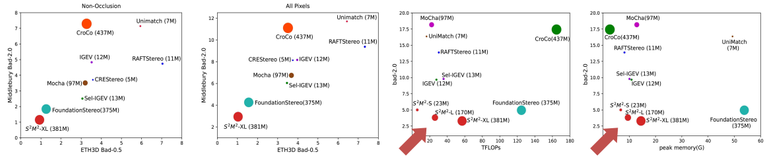
### Ablation Study with Our Synthetic Benchmark

| Model Configuration | | | Training Strategy | Disparity Metrics | | Auxiliary Metrics | |
|---|---|---|---|---|---|---|---|
| Ch | Ntr | AGFL | PMC Loss | EPE | Bad-2.0 | Occ AP | Conf AP |
| 128 | 1 | No | No | 0.489 | 1.974 | 0.973 | 0.977 |
| 128 | 1 | Yes | No | 0.496 | 1.601 | 0.973 | 0.978 |
| 128 | 1 | Yes | Yes | 0.451 | 1.490 | 0.974 | 0.978 |
| 128 | 3 | Yes | Yes | 0.388 | 1.281 | 0.974 | 0.978 |
| 256 | 1 | Yes | Yes | 0.320 | 0.928 | 0.975 | 0.979 |
| 256 | 3 | Yes | No | 0.299 | 0.810 | 0.976 | 0.980 |
| 256 | 3 | Yes | Yes | 0.254 | 0.750 | 0.976 | 0.980 |

| Iterations | GFLOPs | Disparity Metrics | | Auxiliary Metrics | |
|---|---|---|---|---|---|
| | | EPE | Bad-2.0 | Occ AP | Conf AP |
| 0 | 305 | 0.666 | 2.768 | 0.961 | 0.977 |
| 1 | 427 | 0.470 | 1.658 | 0.974 | 0.977 |
| 2 | 537 | 0.479 | 1.526 | 0.974 | 0.977 |
| 3 | 647 | 0.451 | 1.490 | 0.974 | 0.978 |

## Benchmark & Model Scalability




| Model | Non-Occlusion | | | | | | All Pixels | | | | | | Mem(G) | TFLOPs | Param(M) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | EPE | RMS | Bad-0.5 | Bad-1.0 | Bad-2.0 | Bad-4.0 | EPE | RMS | Bad-0.5 | Bad-1.0 | Bad-2.0 | Bad-4.0 | | | |
| CoEx [2] | 20.17 | 63.95 | 38.78 | 27.82 | 21.06 | 16.75 | 37.55 | 103.82 | 46.95 | 36.78 | 29.85 | 24.87 | 31.31 | 0.71 | 2.72 |
| BGNet+ [41] | 14.04 | 42.84 | 43.97 | 29.55 | 20.63 | 14.82 | 30.31 | 88 | 50.98 | 37.85 | 29.11 | 22.87 | 22.49 | 2.42 | 5.31 |
| RAFT [20] | 20.41 | 48.12 | 24.41 | 17.71 | 13.87 | 11.32 | 33.07 | 76.51 | 33.31 | 26.43 | 21.84 | 18.31 | 8.18 | 30.35 | 11.11 |
| fastACV [43] | 15.09 | 48.39 | 36.73 | 24.77 | 17.69 | 13.31 | 32.50 | 93.99 | 45.57 | 34.61 | 27.34 | 22.10 | 8.01 | 0.60 | 3.07 |
| UniMatch [45] | 6.46 | 17.54 | 55.77 | 31.11 | 16.35 | 10.66 | 13.96 | 37.93 | 61.77 | 40.05 | 25.82 | 18.54 | 49.3 | 16.35 | 7.35 |
| CroCo [39] | 15.64 | 35.66 | 33.82 | 23.71 | 17.43 | 13.48 | 24.99 | 57.23 | 40.70 | 30.96 | 24.31 | 19.57 | 2.62 | 164.7 | 437 |
| IGEV [42] | 7.03 | 21.16 | 23.04 | 13.86 | 9.66 | 7.13 | 15.42 | 42.85 | 31.11 | 22.17 | 17.30 | 13.82 | 10.87 | 27.63 | 12.59 |
| MC [11] | 35.56 | 73.77 | 32.48 | 26.72 | 23.08 | 20.56 | 56.07 | 112.67 | 41.12 | 35.47 | 31.46 | 28.29 | 7.73 | 899 | 24.62 |
| NMRF [14] | 40.88 | 82.83 | 58.67 | 43.24 | 36.66 | 33.07 | 62.04 | 121.41 | 64.06 | 50.57 | 44.28 | 40.31 | 9.07 | 3.28 | 6.113 |
| Sel-IGEV [37] | 9.23 | 25.92 | 22.37 | 14.09 | 9.79 | 7.36 | 18.23 | 48.79 | 30.06 | 21.82 | 16.86 | 13.49 | 13.4 | 35.77 | 13.14 |
| MoCha [5] | 40.37 | 102.65 | 26.9 | 21.62 | 18.17 | 15.73 | 59.75 | 141.19 | 35.34 | 29.91 | 25.94 | 22.82 | 12.87 | 22.14 | 97.14 |
| S²M²-S (+1Mtr) | 6.45 | 16.36 | 10.10 | 11.46 | 7.63 | 5.39 | 11.39 | 31.06 | 26.94 | 18.93 | 14.15 | 10.78 | 8.17 | 5.76 | 23.91 |
| S²M²-L (+1Mtr) | 3.05[3] | 10.17[3] | 14.92 | 8.72 | 5.58 | 3.74 | 8.21[3] | 25.92 | 23.48 | 16.53 | 12.21 | 9.12 | 11.39 | 25.96 | 170 |
| FoundationStereo [40] (+2Mtr) | 4.60 | 11.27 | 10.96[3] | 7.23[3] | 4.95[3] | 3.49[3] | 8.26 | 23.04[3] | 18.42[3] | 13.98[3] | 10.73[3] | 8.08[3] | 53.78 | 124.65 | 375 |
| S²M²-S (+2Mtr) | 4.82 | 10.83 | 13.13 | 7.62 | 5 | 3.57 | 9.14 | 26.32 | 20.75 | 14.55 | 10.83 | 8.26 | 8.17 | 5.76 | 23.91 |
| S²M²-L (+2Mtr) | 2.99[2] | 8.72[2] | 10.60[2] | 6.00[2] | 3.80[2] | 2.54[2] | 7.20[2] | 22.40[2] | 18.11[2] | 12.72[2] | 9.39[2] | 6.96[2] | 11.39 | 25.96 | 170 |
| S²M²-XL (+2Mtr) | **2.30[1]** | **7.98[1]** | **9.22[1]** | **5.15[1]** | **3.29[1]** | **2.18[1]** | **6.2[1]** | **20.58[1]** | **16.51[1]** | **11.60[1]** | **8.86[1]** | **6.36[1]** | 14.84 | 56.94 | 381 |

### Critical Re-evaluation of the KITTI Benchmark