PERFORMANCE OF OPTICAL FLOW MODELS ON REAL-WORLD OCCLUDED REGIONS

Author: Iris Petre[1] (i.a.petre@student.tudelft.nl)
Supervisor: Sander Gielisse[1]
Responsible Professor: Jan van Gemert[1]
[1]Delft University of Technology

1. INTRODUCTION

- **Optical flow** estimation is the task of predicting apparent motion of objects between image pairs.
- Used for tasks like **robotics, medical applications** and **object detection**.
- **Occlusions**, where parts of an object become temporarily **hidden**, make accurate motion estimation particularly difficult.
- Recent models use **transformers** or **multi-scale reasoning** to improve handling of occlusions.
- **No existing benchmark** focuses specifically on **occluded** regions.
- Reported results often reflect overall model accuracy **without isolating** occlusion-specific performance.
- Pretraining is typically done on **synthetic** data, then fine-tuned on **real-world** datasets like **KITTI**[3].

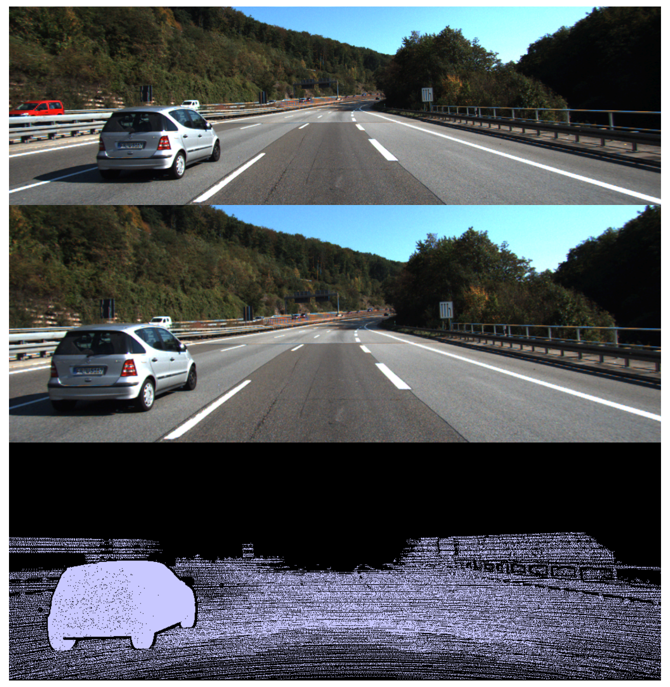

Figure 1. Example scene from KITTI [3] showing limited occlusion coverage. Despite the presence of multiple moving cars, valid flow labels are provided only for the one in the foreground

2. RESEARCH QUESTION

How do state-of-the-art optical flow models perform under real-world occluded regions?

Supporting question: How do the models perform under different types of occluded areas?

3. ANNOTATION PROCESS

Models to evaluate: FlowFormer++[1] and CCMR[2]
Occlusion types:
- **self-occlusion:** a part of an object becomes invisible in the second frame due to perspective transformation
- **inter-object occlusion:** an object is partially hidden by another object in the second frame
- **out-of-frame occlusion**: (a part of) the object leaves the scene

Datasets:
- A real-world dataset focused on occluded areas.
- A real-world dataset with non-occluded annotated points to assess the impact of confounders on occlusion performance.

Dataset Creation & Annotation
- Developed a custom tool for pixel-level optical flow annotation in collaboration with the "Real-world Evaluation of Optical Flow" group.
- Implemented occluson-specific support features
- Developed an annotation method for each occlusion type, such as the Line Intersection Method.

Figure 2. Out-of-frame annotation example showing how pixels outside KITTI [1]'s resolution boundary are labelled.

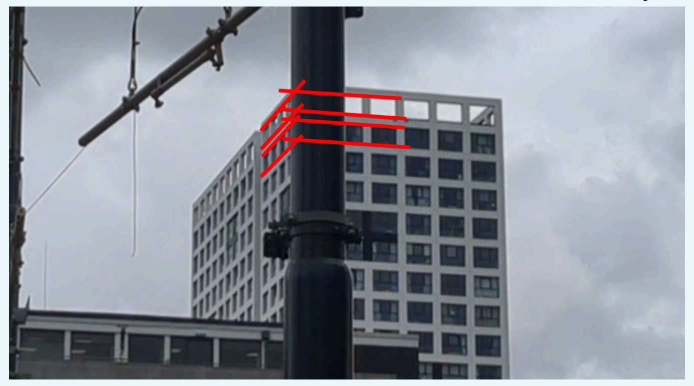

Figure 3: Inter-object occlusion annotation example illustrating how occluded areas are marked at intersections.

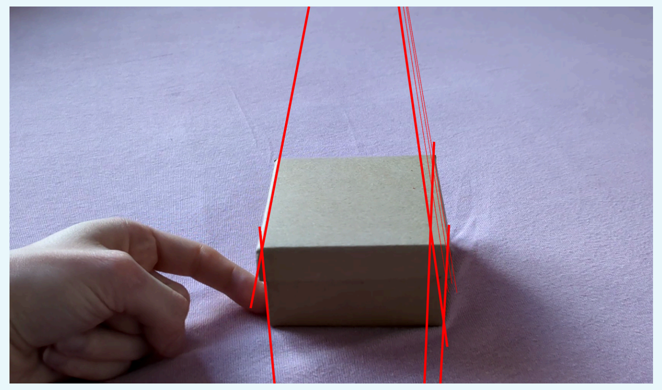

Figure 4: Inter-object occlusion annotation example illustrating how occluded areas are marked at intersections.

4. QUANTITATIVE RESULTS

The **occlusion dataset** contains 22 scenes, 9 outdoor and 13 indoor:
- **95** out-of-frame across 6 scenes
- **94** interobject across 6 scenes
- **95** self-occlusion annotations across 10 scenes

The **non-occluded dataset** uses the same frames, with 106 annotated points.

Annotation accuracy considered: **1-2** pixels for non-occluded and out-of-frame occlusions. **2-3** pixel error margin for inter-object and self-occlusion cases

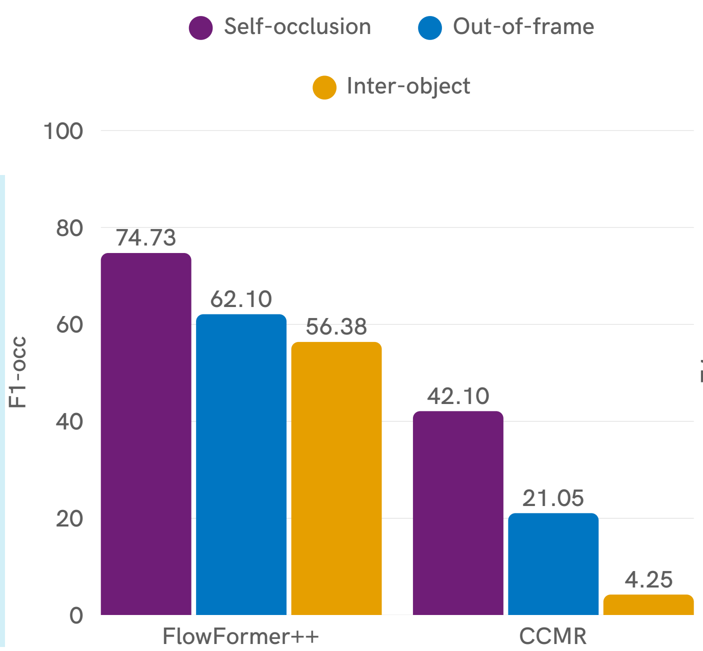

Figure 5: Performance of the models broken down by occlusion type, over all occluded pixels (F1-occ).

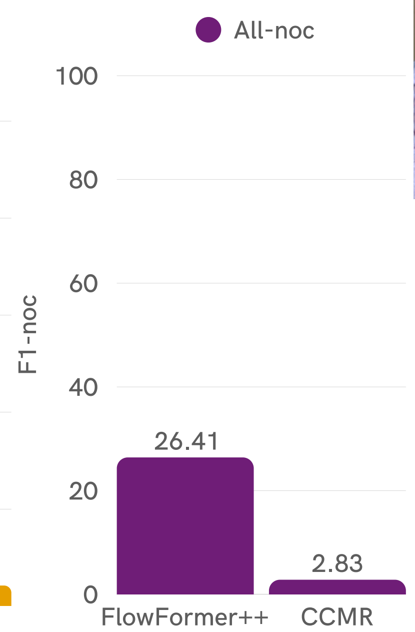

Figure 6: Performance on non-occluded pixels (F1-noc) with FlowFormer++ [1] underperforming significantly

5. QUALITATIVE RESULTS

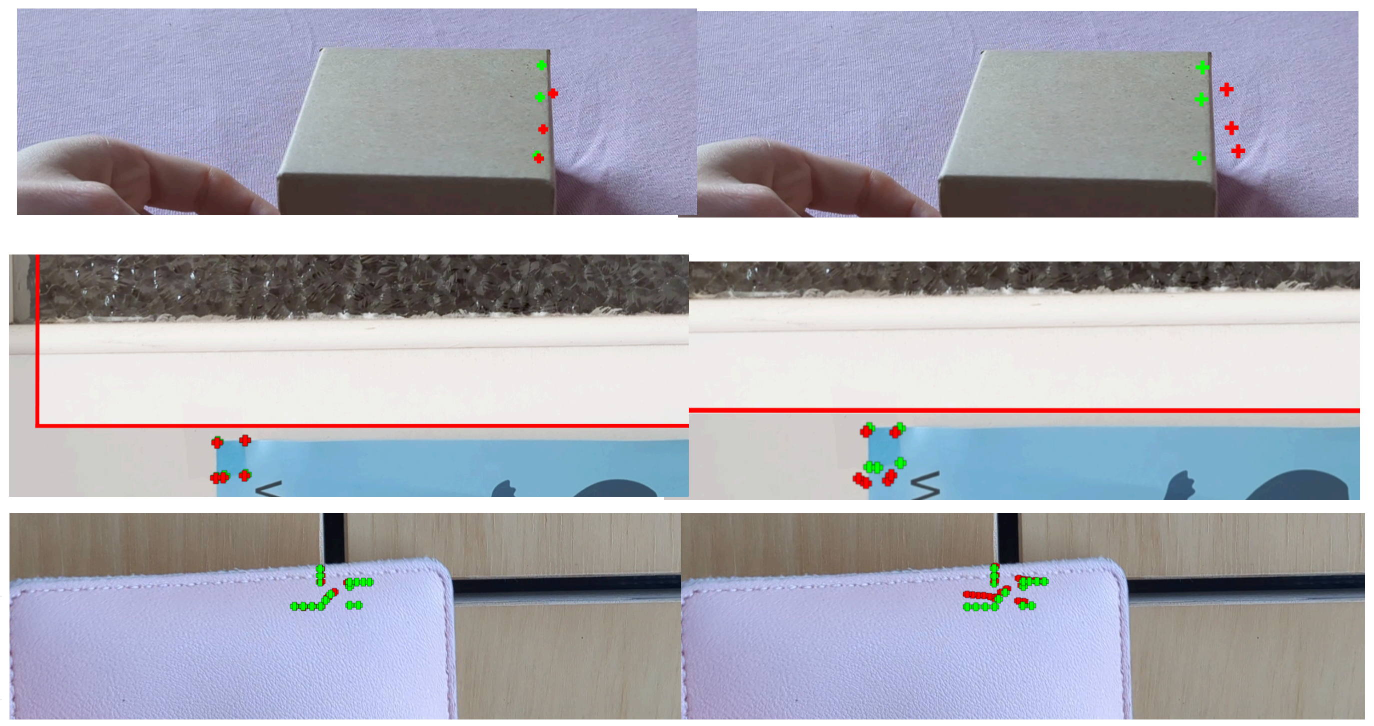

Figure 7. Predicted points (red) vs. ground truth (green) for CCMR [2] (left) and FlowFormer++ [1] (right), across self-occlusion, out-of-frame, and inter-object cases (top to bottom). The middle figures show KITTI's resolution boundary. Results show CCMR's predictions are generally closer to the ground truth than those of FlowFormer++[1].

- Self-occlusions: most challenging likey due to **parallax** and **perspective transformation**
- Out-of-frame occlusions: performance depends on context, strong visual cues help; otherwise, models may **hallucinate** flow.
- Inter-object occlusions: the easiest, likely due to **motion continuity** and because both objects remain **in frame**

6. CONCLUSIONS

- FlowFormer++[1] and CCMR[2] still **struggle** with occlusions, especially with **self-occlusions** being the most challenging, likely due to **rotation** and strong **perspective transformations**.
- **Inter-object occlusions** seem to be the **easiest** for both models to estimate the flow
- **FlowFormer++**[1] seems to **struggle** on scenes with **both** camera and object motion.
- CCMR[2] **outperforms** FlowFormer++[1] significantly, showing greater robustness in both **occluded** and **non-occluded** regions.

7. LIMITATIONS AND FUTURE WORK

- Dataset focuses on occlusion evaluation, but real-world **confounders still affect** performance. However, removing confounders would lead to an **unrealistic and overly simplified dataset**.
- Bias toward **harder** cases because of edge annotations
- No **formal** error margin or multi-annotator validation currently in place.
- **Expand** dataset to include more diverse scenes.
- Improve annotation tool with **semi-automated** features for better efficiency and consistency.

REFERENCES

[1] Xiaoyu Shi et al. FlowFormer++: Masked Cost Volume Autoencoding for Pretraining Optical Flow Estimation. 2023. arXiv: 2303.01237 [cs.CV]. url: https://arxiv.org/abs/2303.01237.
[2] Azin Jahedi et al. CCMR: High Resolution Optical Flow Estimation via Coarse-to-Fine Context-Guided Motion Reasoning. 2023. arXiv: 2311.02661 [cs.CV]. url: https://arxiv.org/abs/2311.02661.
[3] M. Menze and A. Geiger. Object scene flow for autonomous vehicles. In CVPR, pages 3061–3070, 2015.