We thank the reviewers for their positive feedback: well written paper (**R1**,**R2**,**R3**,**R4**) with clear structure and pictures (**R2**,**R3**), interesting method (**R1**,**R3**), useful contribution (**R2**), novelty of the method (**R3**) with well designed losses (**R2**,**R3**), extensive and thoughtful experiments (**R1**,**R2**), superior performance (**R3**) and promising results (**R1**,**R4**).

[R1-Q1] Intuition about objectives: In Sec. 3.5, we aim to provide intuition about our query frame objective Eq. (6) by relating it to classical optical flow methods in an extensive discussion (see L208-221). Regarding Sec. 3.4, we are not aware of similar attempts and thus included an extended description (L150-175) along with visualizations (Fig. 1,3).

[R1-Q2a] Robustness to appearance and geometric variations: As the reviewer points out, the value of additional reference frame information (L103-112) is not as pronounced in semantic matching, since images depict different scenes and object instances. As suggested by the reviewer, we nevertheless evaluate GOCor (without any retraining) for dense semantic matching on the TSS [Taniai, 2016] dataset in Tab. 1. In fact, our GLU-Net-GOCor sets a new state-of-the-art on this dataset, even outperforming Semantic-GLU-Net [49]. Moreover, the results for increasing view-point changes on ETH3D [43] (Fig. 4), indicate that GOCor better copes with large appearance and geometric variations.

Table 1: PCK [%] on TSS.

	FGD3Car	**JODS**	**PASCAL**	**All**
GLU-Net [49]	93.2	73.3	71.1	79.2
Semantic-GLU-Net [49]	94.4	75.5	78.3	82.8
GLU-Net-GOCor	**95.0**	**78.9**	**81.3**	**85.1**

[R1-Q2b] Results without coarse-to-fine: We perform a preliminary experiment by computing the flow directly from the global correlation through an argmax operation. Compared to feature correlation, our GOCor achieves 8.0% and 13.0% better EPE on HPatches [3] and KITTI-2015 [13] respectively. Importantly, the correspondence volume generated by GOCor is also much more discriminative (Fig. 1), greatly enhancing the results of correspondence networks.

[R1-Q3] Using cross entropy in Eq. (5): While our objective function and optimization module could be formulated with cross-entropy instead, our squared loss allows the use of Gauss-Newton for efficient optimization and flexibility through learned parametrization. We will consider this interesting suggestion for future work.

[R1-Q4] Additional references: We thank the reviewer for the references and will include them in the paper.

[R2-Q1] Interchanging query and reference features: Prior to submission, we experimented with also interchanging the two frames at the global level, and then fusing the two resulting GOCor correspondence volumes. However, we only observed marginal improvements. E.g., on KITTI-2015 it obtains an EPE of 11.07 and an F1 of 54.68% compared to 10.97 EPE and 55.62% F1 for the baseline 'BaseNet $L_r + L_q$' (suppl. Tab. 7). We will include this experiment.

[R2-Q2] Nature of optimizer (L231-242): It is an online process performed at every forward pass of the network.

[R2-Q3, R3, R4-Q1] Computational complexity: We perform a detailed analysis of the run-time for varying number of optimizer iterations in suppl. Tab. 2 (Sec. E.1) and Tab. 8, which we will move to the main paper. While our GOCor has an impact in run time, we believe that it is small compared to the improvement in performance brought by our module. In suppl. Sec. E.2 we discuss and suggest a good speed-accuracy trade-off.

[R3-Q1] Properties of the regularizer (Sec. 3.5): With our claim in L219 we mean that our formulation in Eq. (6) is capable of learning filters R_θ that can enforce local uniqueness (will be clarified). While we did not claim that it actually does, this is a very interesting question that is, however, difficult to verify empirically. In practice, we often observe that the query objective also has a 'peak-enhancing' effect, as shown in Fig. 1 below.

[R3-Q2] Smoothness constraints on the flow field: While this is an interesting point, it is very difficult to compare the two strategies in practice, since the flow in our approach is predicted with a deep CNN. As a result, a regularization loss on the flow itself cannot easily be embedded into our objective. We will nevertheless consider this for future study.

[R3-Q3] Performance on occlusion data: As shown in suppl. Tab. 5, in occluded regions ("EPE unmatched") of the Sintel test set, GOCor provides relative improvements of 6.25% and 2.16% on the clean and final pass respectively. On KITTI-2015, GOCor improves the performance of PWC-Net and GLU-Net in occluded regions, as shown in Tab. 2. Moreover, we did not observe noticeable blurring at occlusion boundaries.

Table 2: AEPE/F1 [%] on KITTI-2015.

	Not occluded	**Occluded**	**All**
GLU-Net	4.67 / 27.83	21.95 / 67.44	7.49 / 33.83
GLU-Net-GOCor	**4.22 / 22.03**	**19.07 / 58.61**	**6.68 / 27.57**
PWC-Net	5.40 / 25.16	34.39 / 78.58	10.81 / 32.75
PWC-Net-GOCor	**5.02 / 23.53**	**34.06 / 77.84**	**10.33 / 30.53**

[R4-Q2] Clarification of the design and name of the query frame objective: We call Sec. 3.4 and 3.5 the reference and query frame objectives since they are evaluated based on the correspondence volume predicted on the reference and query frame respectively. In Eq. (6), the convolutional kernel R_θ is applied to the correspondence volume $C(w, f^q)$ between our filter map w and the query feature map f^q. Note that it is the filter map w that is optimized using Eq. (5) and (6) at each forward pass, while the kernel R_θ (L204) is learnt, along with all other network parameters, by the SGD-based minimization of the same final network training loss used in the GLU-Net and PWC-Net baselines.

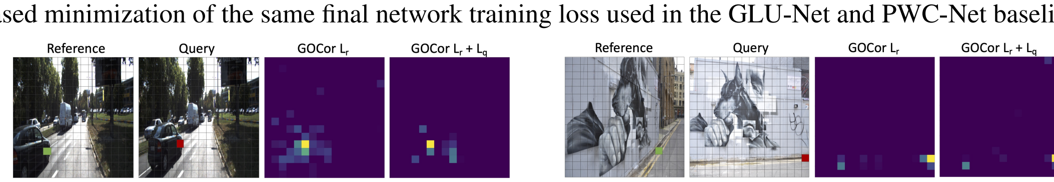

Figure 1: Visualization of the matching confidences computed between the indicated location (green) in the reference image and all locations of the query image. We compare with and without utilizing the query frame objective L_q.