RAFT

Under the guidance of:

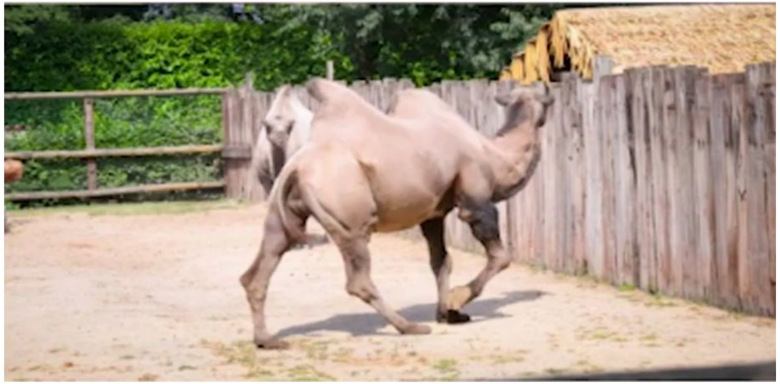
DR. DINESH NAYAK

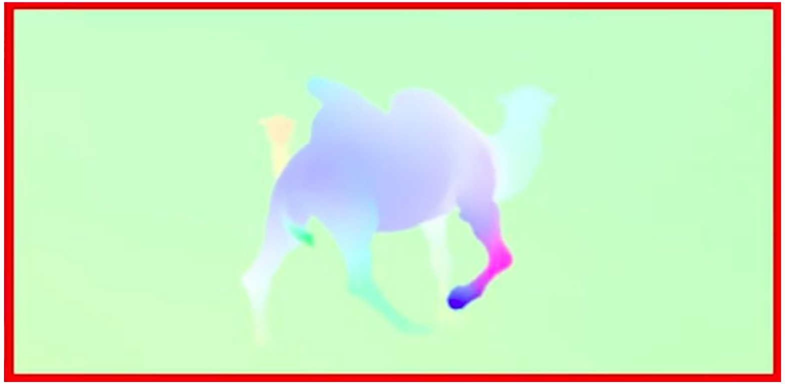
Presented by:

MOHD ASIF KHAN
(202IT013)
PRAFUL KUMAR
(202IT020)

Agenda

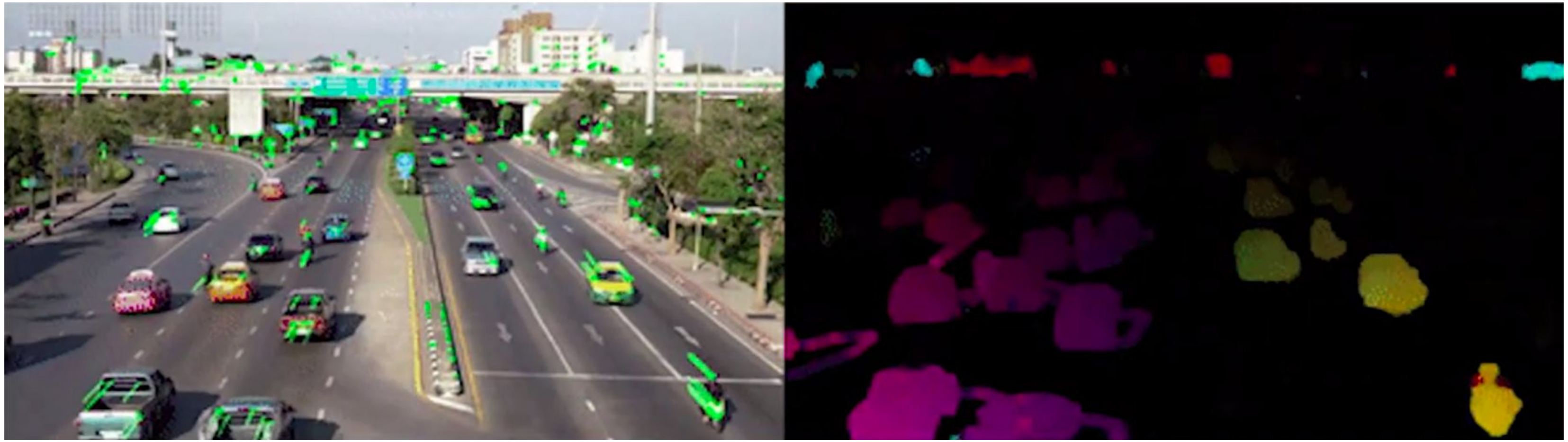
- Optical flow
- Abstract
- Related work
- Approach
- Methodology
- Dataset
- Experiments and Results
- Conclusion
- Future work
- References





Optical Flow

Optical flow is the task of estimating per-pixel motion between video frames. It is a long-standing vision problem that remains unsolved. The best systems are limited by difficulties including fast-moving objects, occlusions, motion blur, and textureless surfaces.



Abstract

- Recurrent All-Pairs Field Transforms (RAFT), a new deep network architecture for optical flow. RAFT extracts per- pixel features, builds multi-scale 4D correlation volumes for all pairs of pixels, and iteratively updates a flow field through a recurrent unit that performs lookups on the correlation volumes.
- RAFT achieves state- of-the-art performance. On KITTI, RAFT achieves an F1-all error of 5.10%, a 16% error reduction from the best published result (6.10%). On Sintel (final pass), RAFT obtains an end-point-error of 2.855 pixels, a 30% error reduction from the best published result (4.098 pixels).
- In addition, RAFT has strong cross-dataset generalization as well as high efficiency in inference time, training speed, and parameter count.

Related work

- Direct Flow prediction:
    - Supervised task
    - Direct prediction using encoder decoder
    - Usually multiple encoders and decoders are stacked to have a coarse refinement.
    - FlowNet
- Iterative refinement of optical flow:
    - Using a subnetwork to iteratively update the residual of optical flow
    - Limited by subnetwork size.
    - A typical work(Iterative residual refinement) on this iterate 5 times, RAFT uses small module and iterate over 100 times.

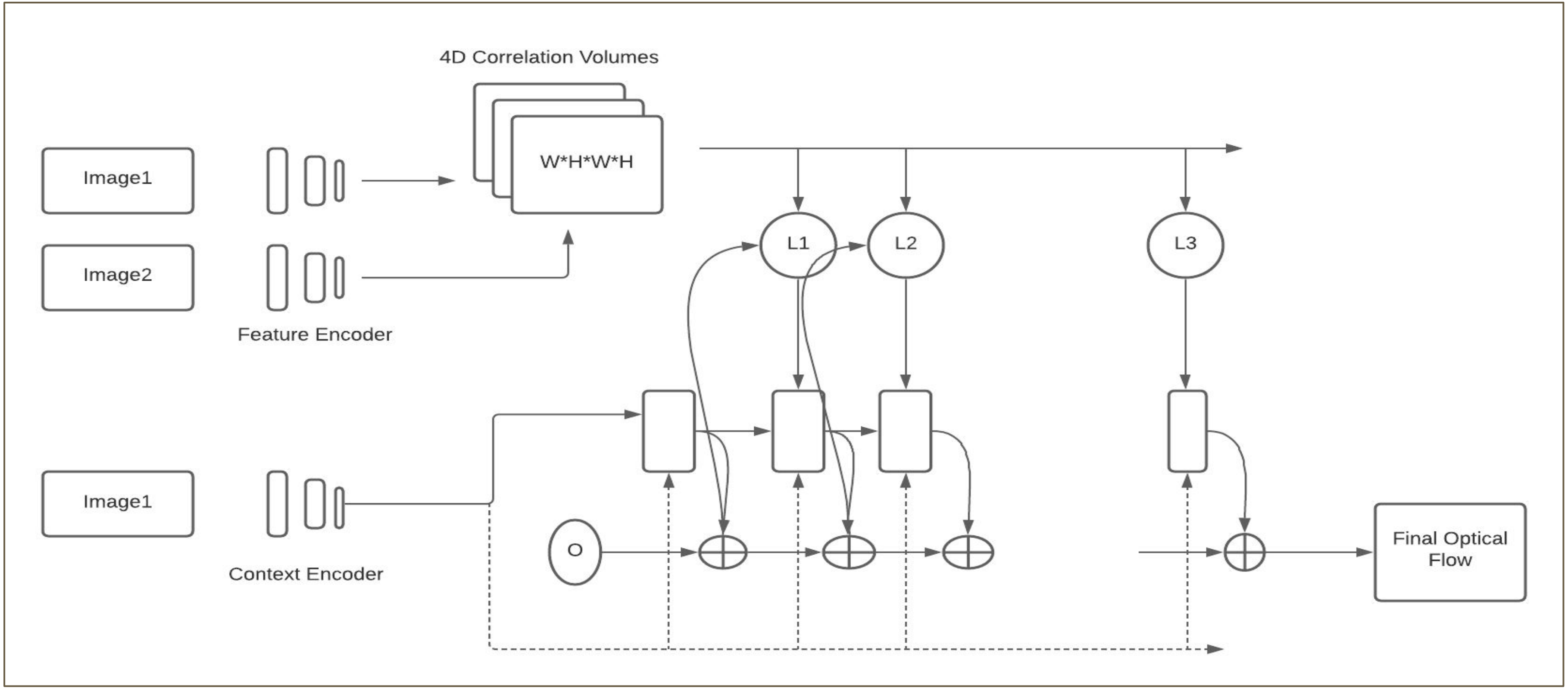
- Learning to optimize
  - Network can be used to predict the input parameter of optimization problem, and therefore can be trained end to end as long as the optimization process is differentiable.
  - RAFT can be viewed as Learning to optimize: It has a iterative update module, which resembles the first order optimization algorithm.

Approach

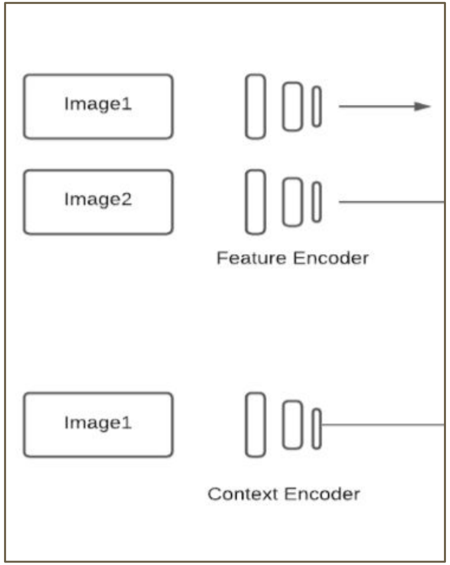
RAFT consists of three main components:

1. a **feature encoder** that extracts a feature vector for each pixel
2. a **correlation layer** that produces a 4D correlation volume for all pairs of pixels, with subsequent pooling to produce lower resolution volumes
3. a **recurrent GRU-based update operator** that retrieves values from the correlation volumes and iteratively updates a flow field initialized at zero.

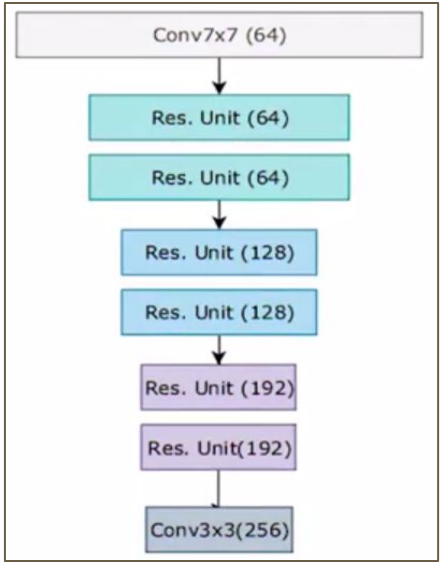
Methodology



Feature extraction

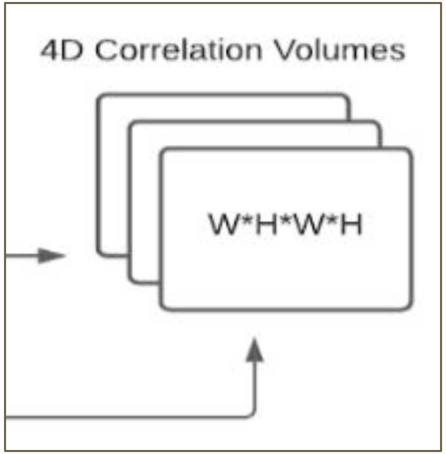


- Here we have separate blocks for computing **context encoder** and **feature encoder**.
- 6 residual blocks:
  - 2 x ½ resolution
  - 2 x ¼ resolution
  - 2x ⅛ resolution



Computing visual similarity

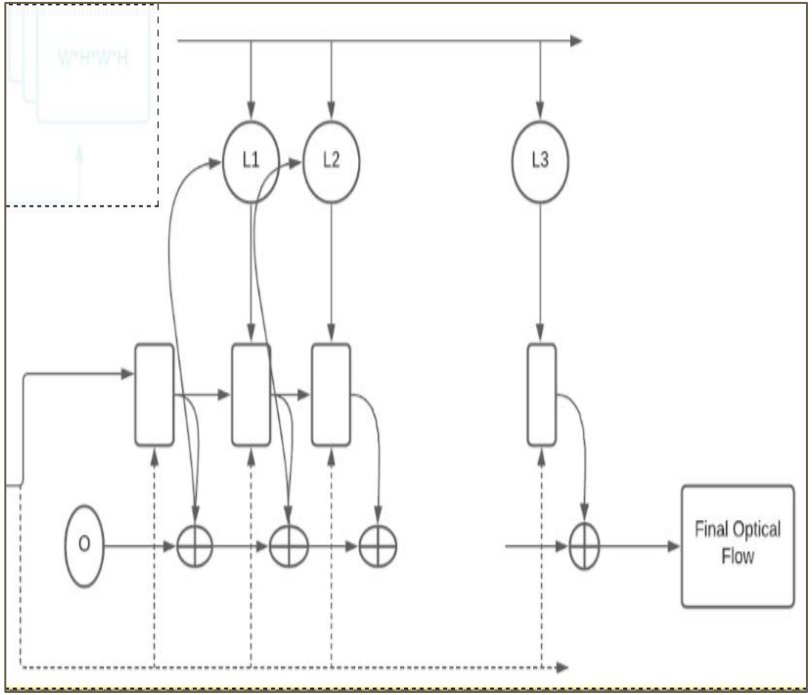
- Correlation matrices store the visual similarity between both the input images.
- Computed using dot product of feature vectors as computed in step 1.
- We calculate the 4D correlation volume WxHxWxH.
- Each pixel in image 1 can produce the response map of image 2.
- Finally, 2x2 pooling applied 3 times at different scales.
- $C \in R^{WxHxWxH}$



Correlation Lookup

- Given the current estimates (f1, f2), x = (u,v) in image 1.
- The correspondence pixel in image 2 is x' = (u+f1(u), v+f2(v)).
- Define the local grid within the 1-norm ball.
- $N(x')_r = \{x' + dx \mid dx \in Z^2, ||dx||_1 <= r\}$
- End up with a shape: (H, W, LEN($N(x')_r$),4)
- Done for all pixels and result is concatenated.

Iterative update



- Given an initial estimate of flow.
- Iteratively refine the flow f by correlation lookup and update operator based on previous estimated flow.
- Update block is based on gated activation unit.
- Loss function: for all the outputs {f1,...,fn}

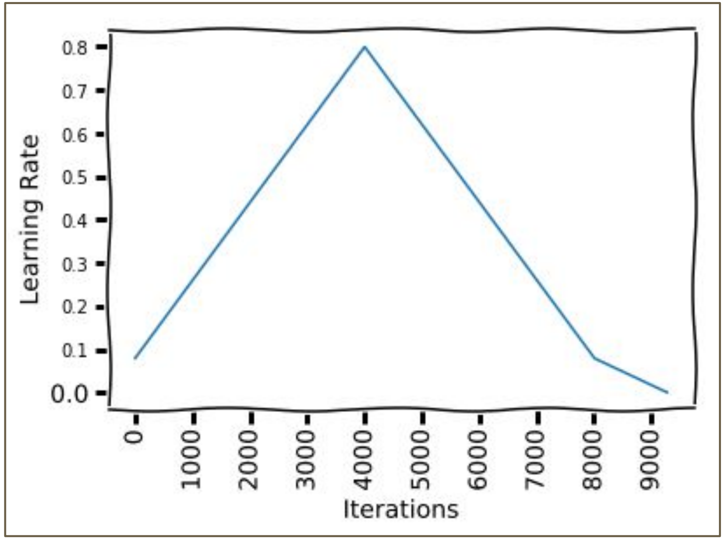
  $L = \sum(Y^{n-i}||f_{gt}-f_i||_1)$.

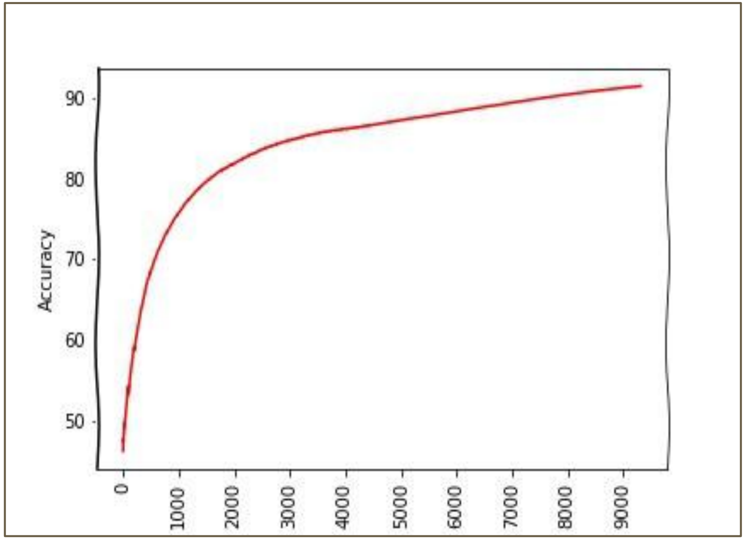
Dataset

- **MPI Sintel**
- open source animated short film.
- most widely used datasets for training and evaluating optical flow algorithms.
- extremely challenging and current methods have still not fully "solved" the problem of flow estimation for Sintel.
- designed to encourage research on long-range motion, motion blur, multi-frame analysis, non-rigid motion.
- contains flow fields, motion boundaries, unmatched regions, and image sequences.
- also provide ground truth depth, stereo, and camera motions.

Training

- **One Cycle Learning Policy**

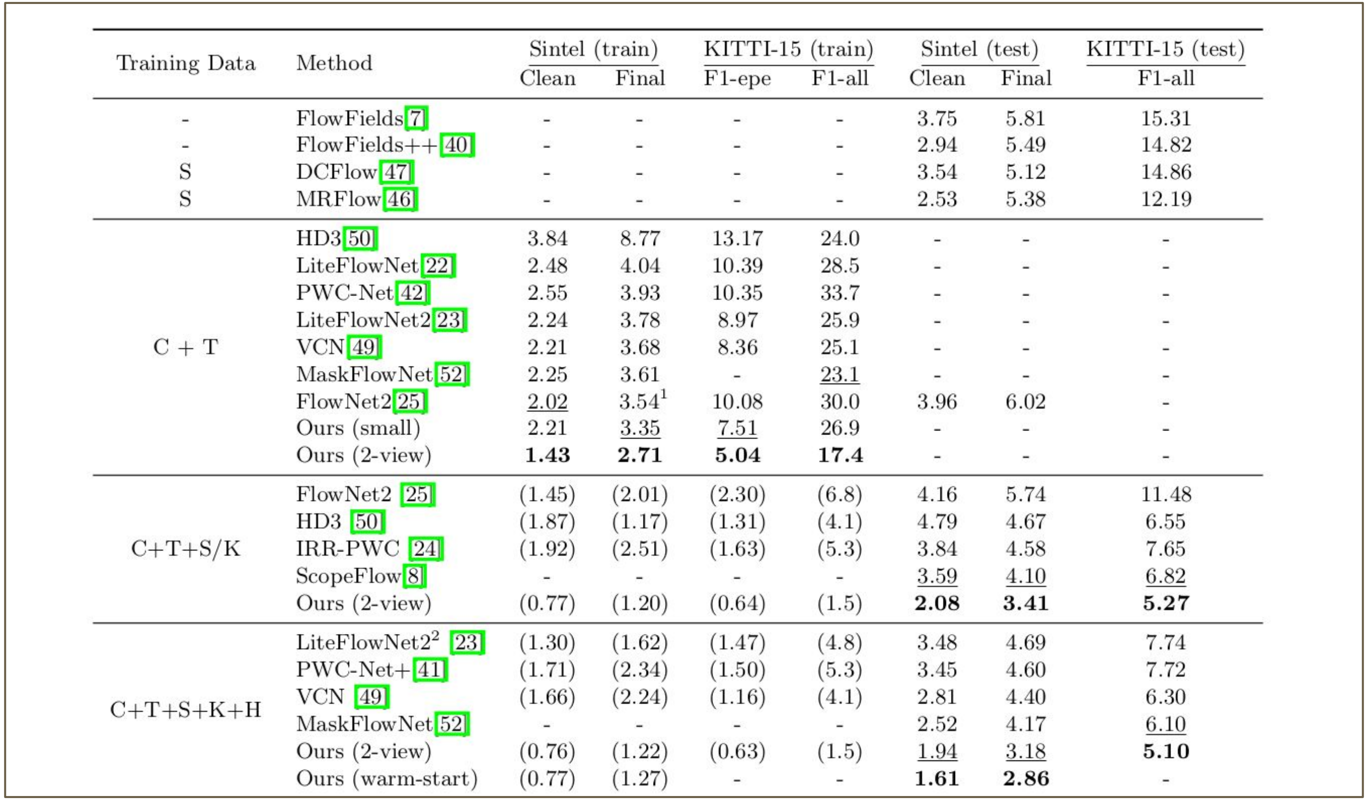




One Cycle Learning Policy

- **Problem with static learning rate** : takes much training to get optimal learning rate.
- **Solution** : Cyclic learning rates given in 2018 paper by “A DISCIPLINED APPROACH TO NEURAL NETWORK HYPER-PARAMETERS: PART 1 – LEARNING RATE, BATCH SIZE, MOMENTUM, AND WEIGHT DECAY”.

Authors Results



Training Data	Method	Sintel (train)		KITTI-15 (train)		Sintel (test)		KITTI-15 (test)
		Clean	Final	F1-epe	F1-all	Clean	Final	F1-all
-	FlowFields [7]	-	-	-	-	3.75	5.81	15.31
-	FlowFields++ [40]	-	-	-	-	2.94	5.49	14.82
S	DCFlow [47]	-	-	-	-	3.54	5.12	14.86
S	MRFlow [46]	-	-	-	-	2.53	5.38	12.19
C + T	HD3 [50]	3.84	8.77	13.17	24.0	-	-	-
	LiteFlowNet [22]	2.48	4.04	10.39	28.5	-	-	-
	PWC-Net [42]	2.55	3.93	10.35	33.7	-	-	-
	LiteFlowNet2 [23]	2.24	3.78	8.97	25.9	-	-	-
	VCN [49]	2.21	3.68	8.36	25.1	-	-	-
	MaskFlowNet [52]	2.25	3.61	-	23.1	-	-	-
	FlowNet2 [25]	2.02	3.54[1]	10.08	30.0	3.96	6.02	-
	Ours (small)	2.21	3.35	7.51	26.9	-	-	-
	Ours (2-view)	**1.43**	**2.71**	**5.04**	**17.4**	-	-	-
C+T+S/K	FlowNet2 [25]	(1.45)	(2.01)	(2.30)	(6.8)	4.16	5.74	11.48
	HD3 [50]	(1.87)	(1.17)	(1.31)	(4.1)	4.79	4.67	6.55
	IRR-PWC [24]	(1.92)	(2.51)	(1.63)	(5.3)	3.84	4.58	7.65
	ScopeFlow [8]	-	-	-	-	3.59	4.10	6.82
	Ours (2-view)	(0.77)	(1.20)	(0.64)	(1.5)	**2.08**	**3.41**	**5.27**
C+T+S+K+H	LiteFlowNet2[2] [23]	(1.30)	(1.62)	(1.47)	(4.8)	3.48	4.69	7.74
	PWC-Net+ [41]	(1.71)	(2.34)	(1.50)	(5.3)	3.45	4.60	7.72
	VCN [49]	(1.66)	(2.24)	(1.16)	(4.1)	2.81	4.40	6.30
	MaskFlowNet [52]	-	-	-	-	2.52	4.17	6.10
	Ours (2-view)	(0.76)	(1.22)	(0.63)	(1.5)	1.94	3.18	**5.10**
	Ours (warm-start)	(0.77)	(1.27)	-	-	**1.61**	**2.86**	-

Experiments 1: Manual Hyperparameter tuning

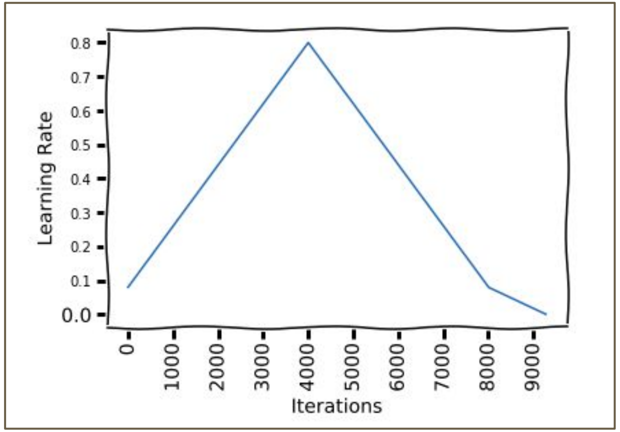
Batch Size	Learning rate	Weight decay	gamma	GRU iterations	steps	Sintel(Clean)	Sintel (final)
10	0.000125	0.00001	0.9	12	5000	4.24	5.03
10	0.000125	0.00001	0.9	12	10000	3.50	4.23
5	0.00125	0.00001	0.9	12	5000	3.76	4.52
10	0.00125	0.00001	0.9	12	10000	2.43	3.16
5	0.0001	0.00001	0.9	12	5000	5.67	6.44
5	0.0001	0.00001	0.9	12	5000	3.83	4.54

Experiments 1: Manual Hyperparameter tuning

Batch Size	Learning rate	Weight decay	gamma	GRU iterations	steps	Sintel(Clean)	Sintel (final)
10	0.000125	0.00001	0.9	12	5000	4.25	5.05
10	0.000125	0.00001	0.9	12	10000	3.38	4.22
13	0.000125	0.00001	0.85	6	5000	5.16	5.19
13	0.000125	0.00001	0.85	6	10000	4.06	4.79

Experiments 2 : One Cycle Learning

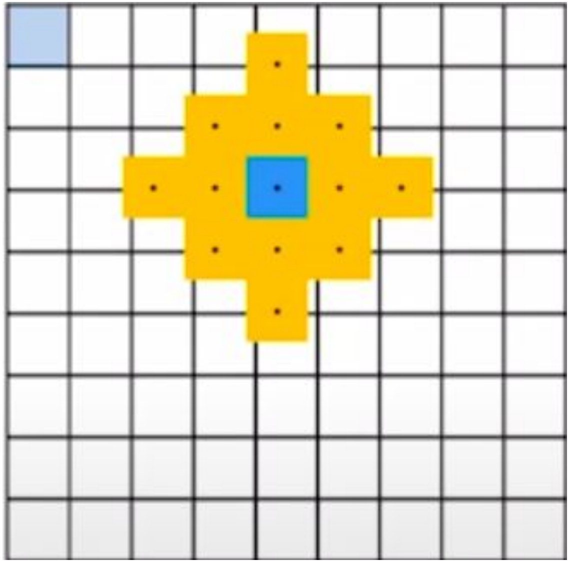
Batch Size	Learning rate	Weight decay	gamma	GRU iterations	steps	cycle_momentum	Sintel(Clean)	Sintel (final)
13	0.000125	0.00001	0.85	6	10000	true	3.02	3.80



Experiments 3 : Transfer Learning

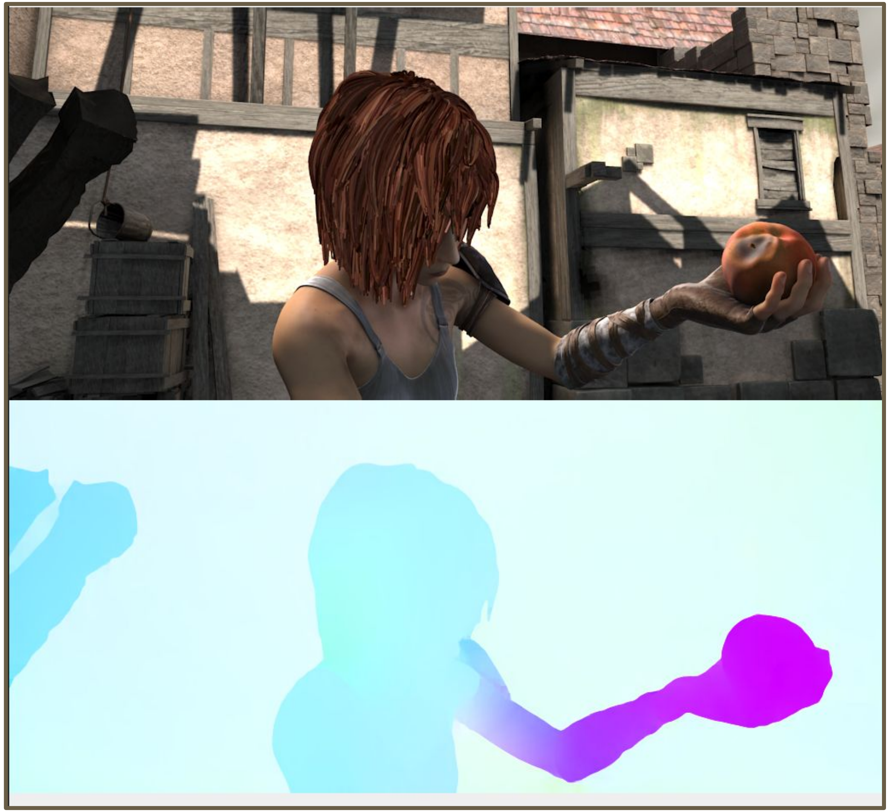
Batch Size	Learning rate	Weight decay	gamma	GRU iterations	steps	Sintel(Clean)	Sintel (final)
13	0.000125	0.00001	0.85	6	5000	4.25	5.05
13	0.000125	0.00001	0.85	6	10000	2.10	3.26
13	0.000125	0.00001	0.85	12	5000	1.81	2.75
13	0.000125	0.00001	0.85	12	10000	1.60	2.34

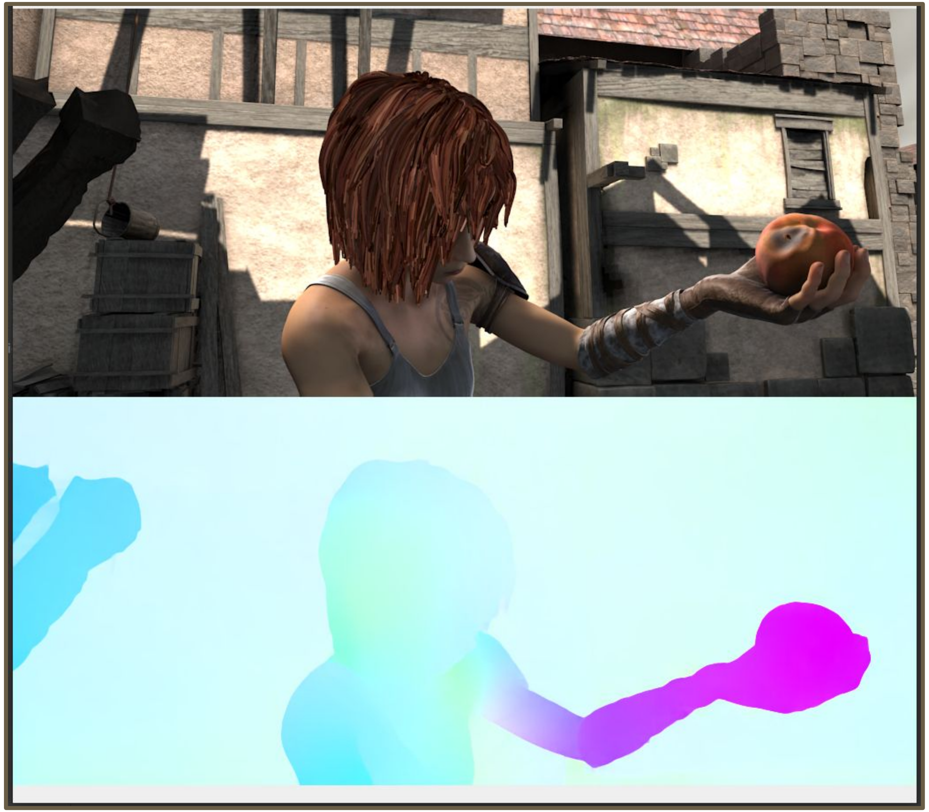
Experiments 4 : Correlation Lookup Radius

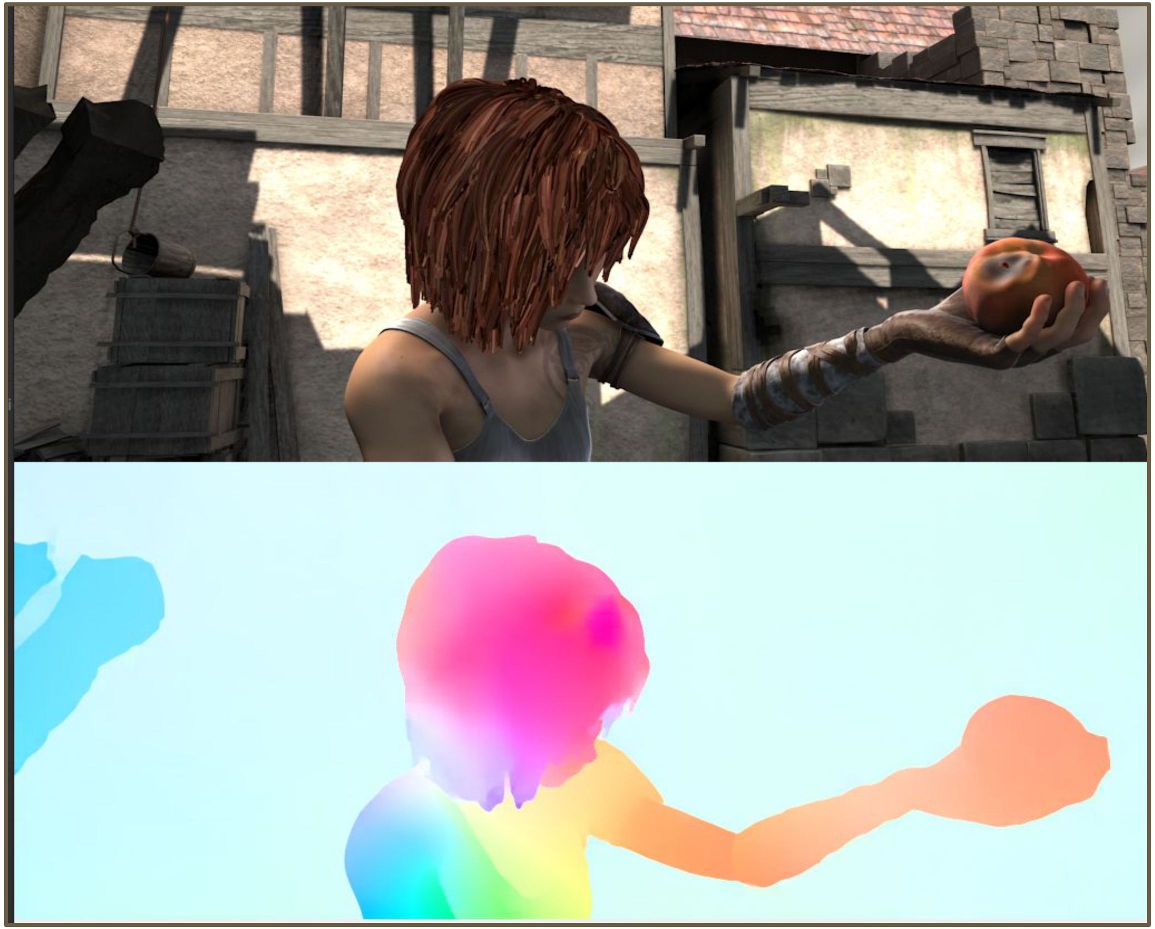


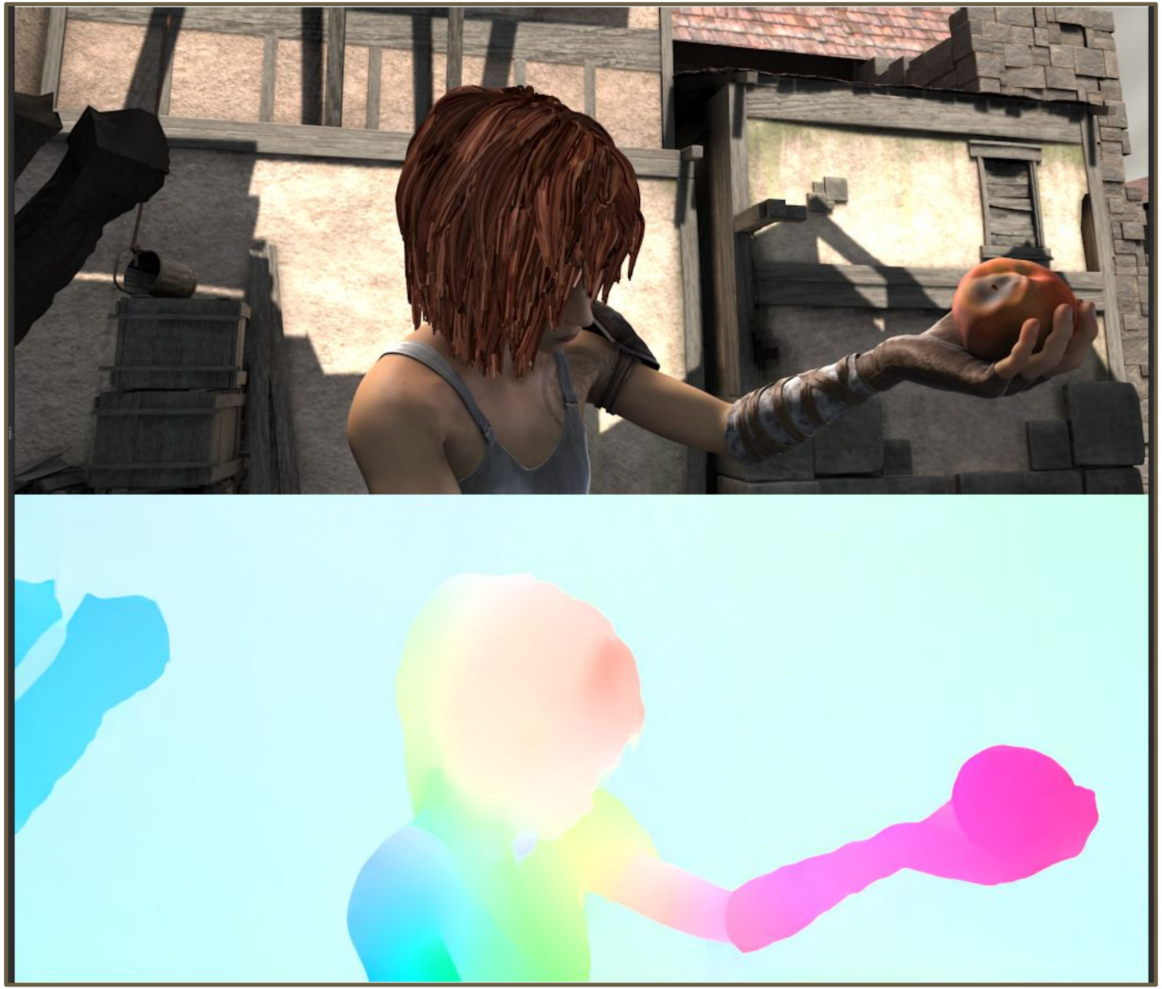
Radius	Sintel(Clean)	Sintel (final)
4	1.60	2.34
3	1.58	2.20
5	1.56	2.29

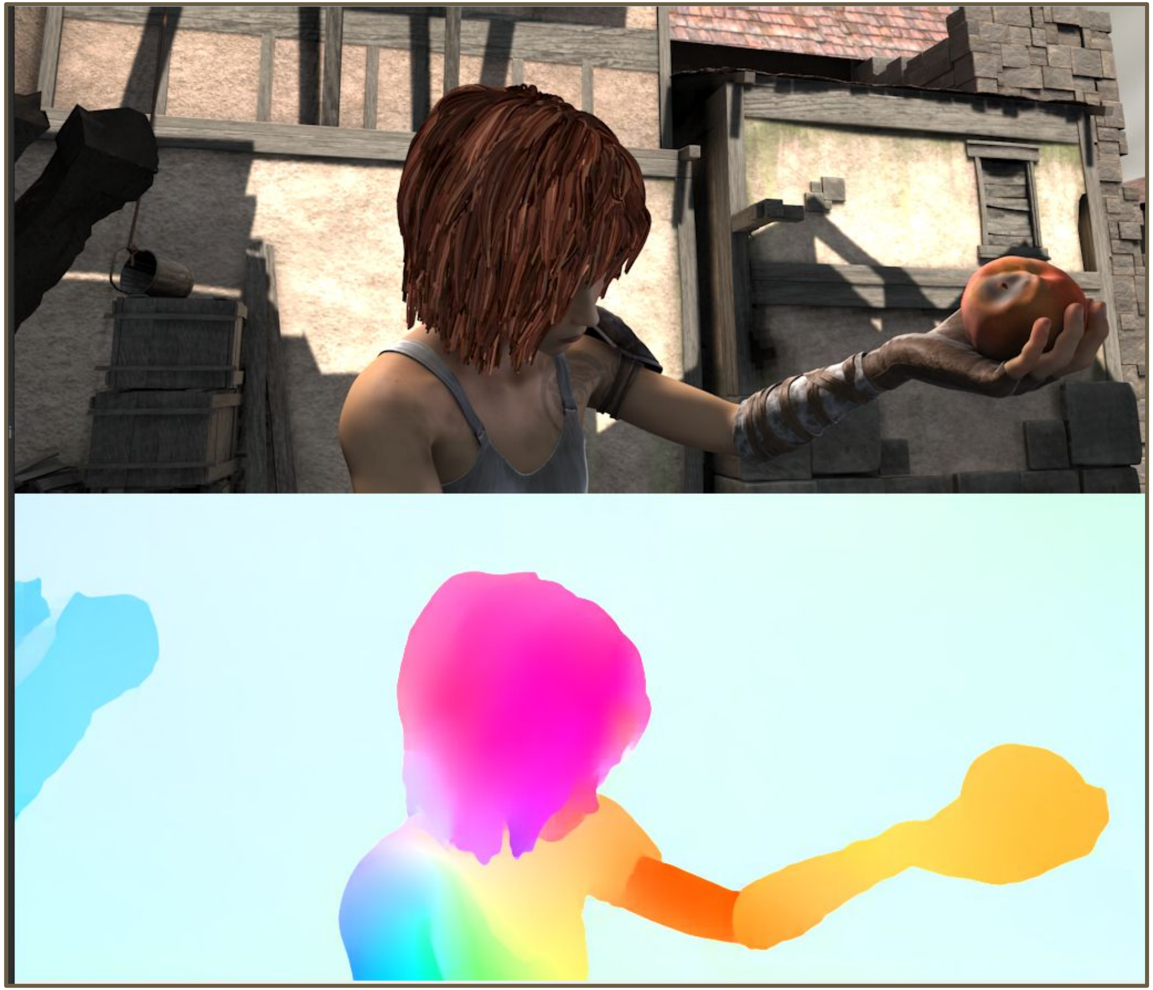
Optical Flow











Conclusion

- Based on the experiments we did on the Sintel dataset we got comparative performance to RAFT original model and better performance to RAFT small model.
- Also the best results were obtained when we used transfer learning on a pretrained model. The model is also trained for less steps than the original number of steps mentioned in RAFT.

Future Work

We would like to improve the architecture further by modifying the correlation pyramid as well as GRU iterations part because that is an iterative process and take some time but essential for good prediction of optical flow using context information from first image. In future we will try to improve that pipeline as well. Also the EPE can be improved further with respect to base RAFT.

References

- Zachary Teed, & Jia Deng. (2020). RAFT: Recurrent All-Pairs Field Transforms for Optical Flow.
- Leslie N. Smith (2018). A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay*. CoRR, abs/1803.09820.*
- https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_video/py_lucas_kanade/py_lucas_kanade.html
- FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks(2018)
- Liteflownet: A lightweight convolutional neural network for optical flow estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition(2018):

Thank You