PROJECT PRESENTATION

RAFT FOR OPTICAL FLOW ESTIMATION

Presented on: 08/11/2024
Adame Dey
Achraf Habib

Supervised by
Dr Rabaa Youssef

TABLE OF CONTENTS

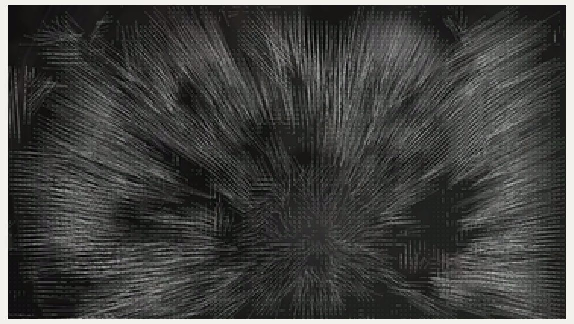
INTRODUCTION

1 INTRODUCTION

problem and context

Optical flow estimation tracks motion between video frames, it’s crucial for applications like autonomous driving, robotics ..

However, it's challenging due to large motions, occlusions, and the need for real-time processing. Traditional methods often struggle with accuracy and speed. RAFT (Recurrent All-Pairs Field Transforms) solves these issues



Let’s explore how ?

Benchmark Performance Examples:

- **KITTI Benchmark:** A popular benchmark for testing vision tasks in self-driving scenarios, KITTI evaluates models on their ability to handle realistic road scenes and large movements. RAFT's strong performance on KITTI shows its capability for real-world driving applications.
- **Sintel Benchmark:** A synthetic benchmark designed to test optical flow models on complex scenes, with varied lighting, textures, and occlusions. RAFT's success on Sintel demonstrates its ability to handle challenging and diverse visual environments, making it suitable for both real-world and artificial scenarios.

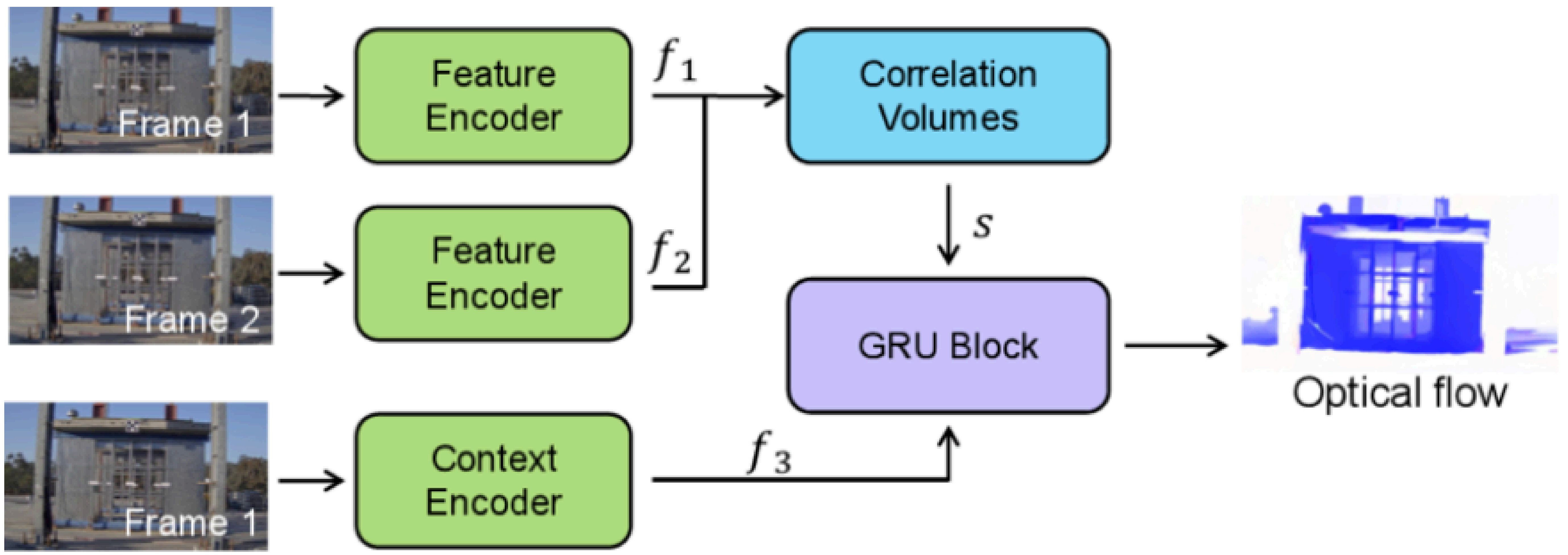
RAFT
VS
FLOWNET2.0

2 RAFT VS FLOWNET 2.0

RAFT	CRITERIA	FLOWNET 2.0
• Uses an all-pairs correlation to compare every pixel in Image A with every pixel in Image B, followed by recurrent flow refinement.	**APPROACH**	• RAFT uses a multi-scale network to iteratively refine optical flow estimates.
• Provides high precision , excelling in large movements, occlusions, and complex scenes.	**Precision**	• While accurate, it tends to struggle more with large displacements and complex scenarios compared to RAFT
• More suitable for applications requiring high accuracy in complex motion tracking (e.g., autonomous driving, AR).	**Use Cases**	• Ideal for tasks requiring reliable general flow estimates, where slight inaccuracies or smaller motion displacements are acceptable.

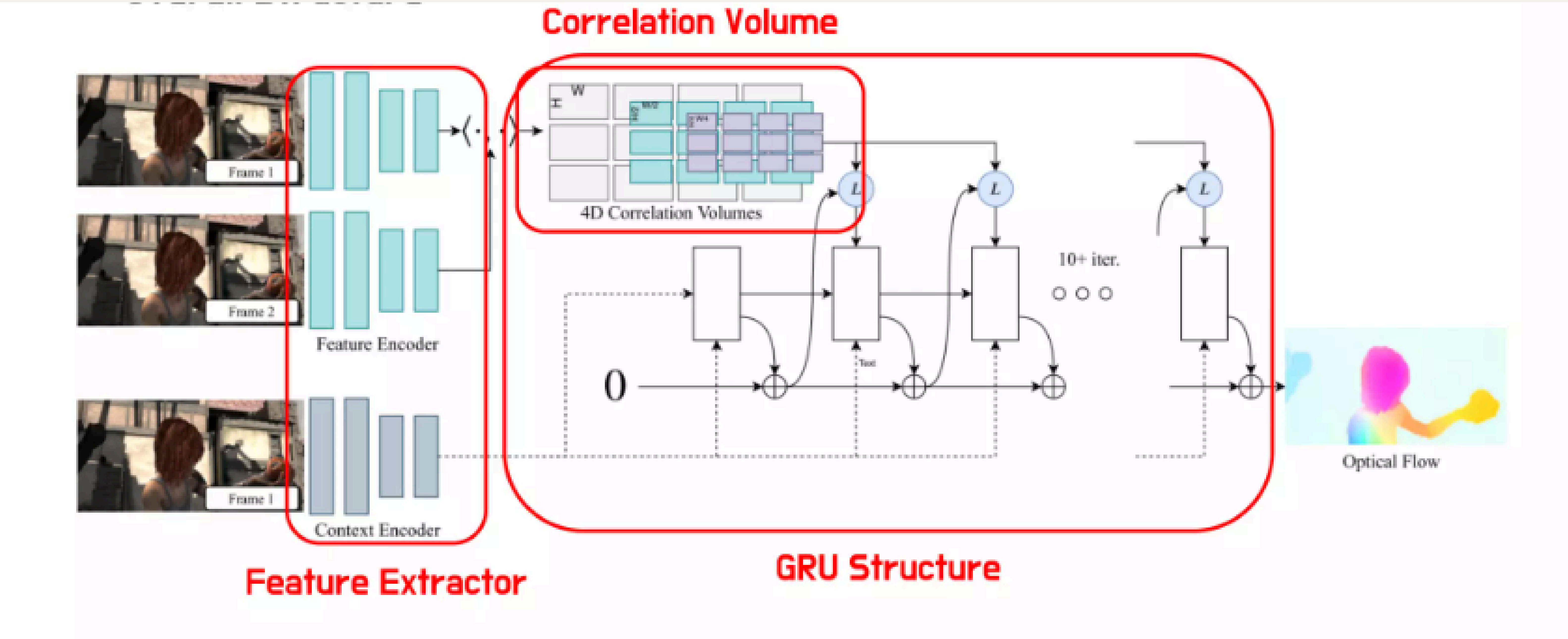
RAFT ARCHITECTURE

2 RAFT ARCHITECTURE



Architecture of RAFT

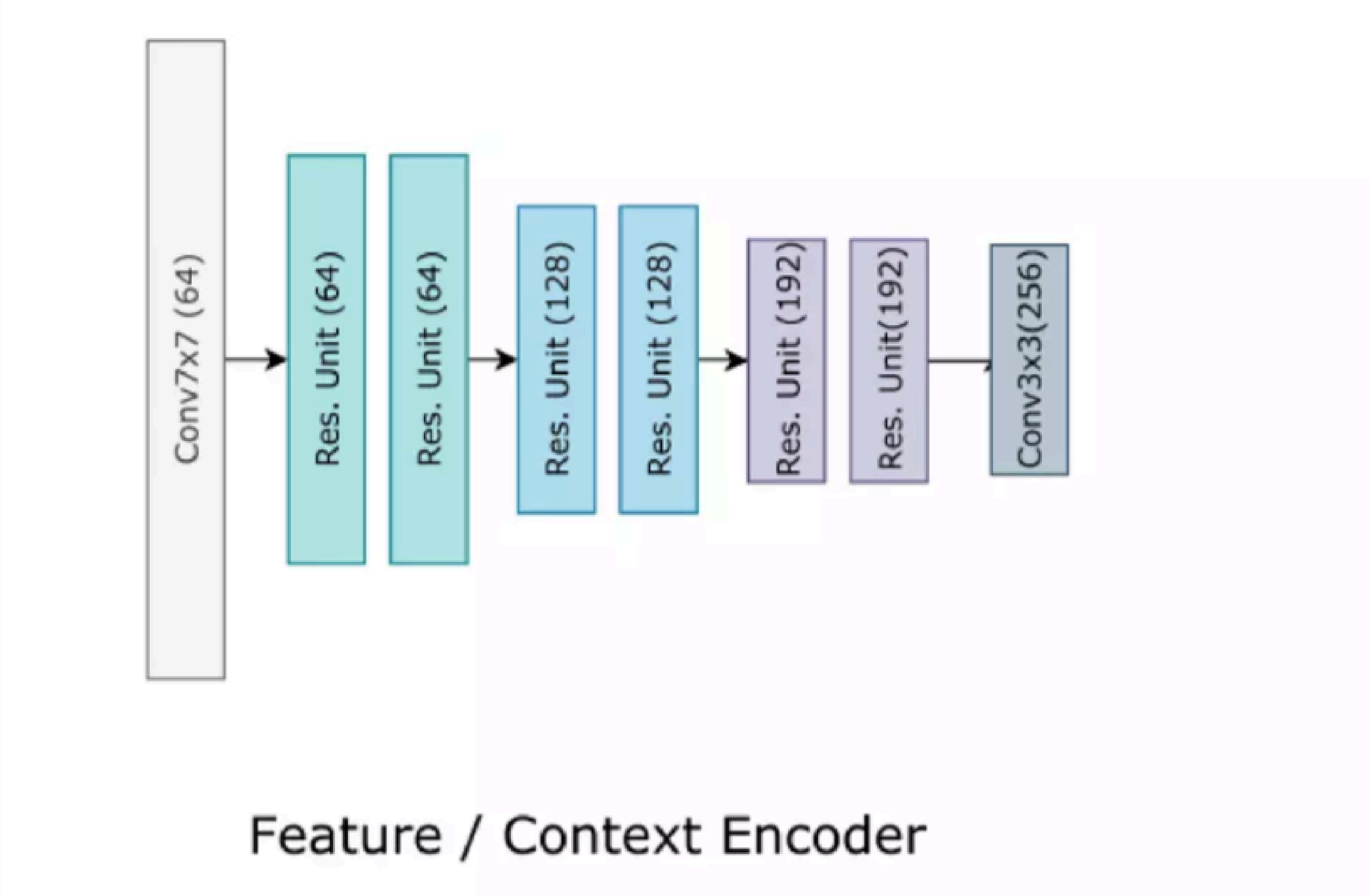
2 RAFT ARCHITECTURE



2 RAFT ARCHITECTURE

FEATURE EXTRACTOR

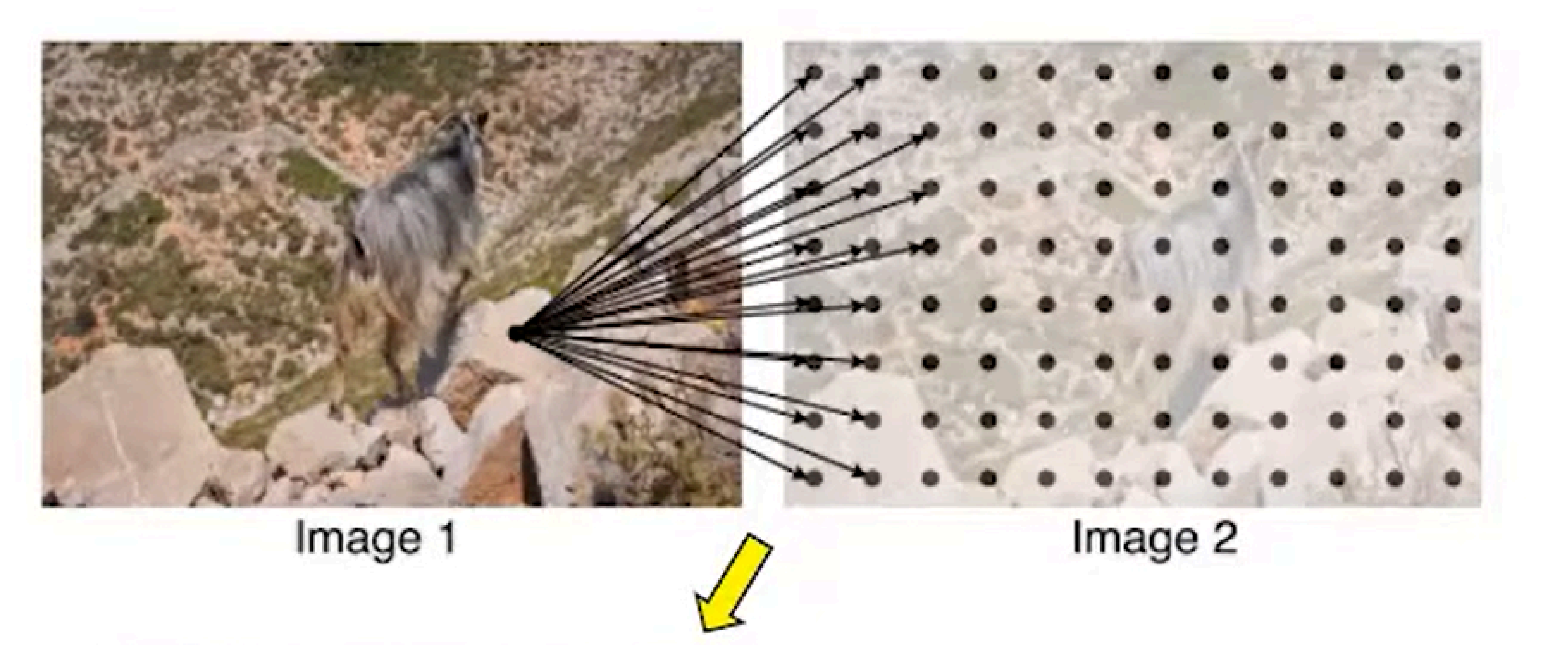
- The process begins with two images, called Image A and Image B.
- These images are fed into a feature extractor network.
- The feature extractor network is a convolutional neural network (CNN).
- The CNN converts each image into a feature map.
- Each feature map is a compressed representation of the image.
- In this feature map, each pixel contains detailed information about its local context.

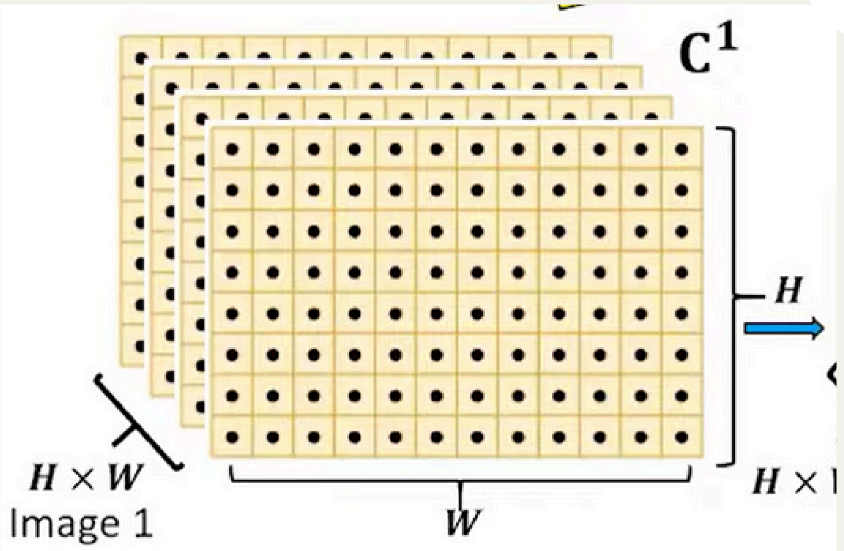


2 RAFT ARCHITECTURE

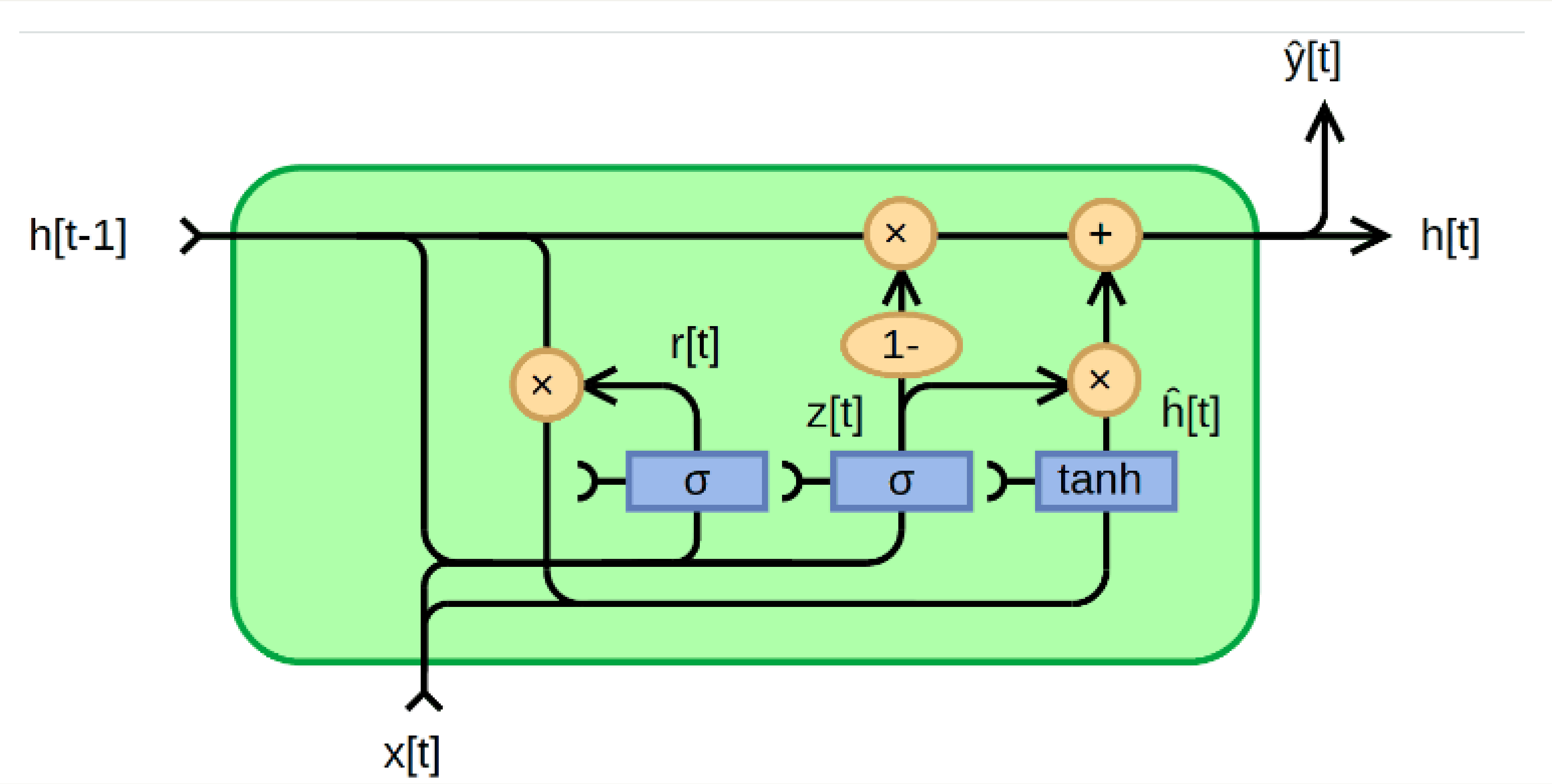
ALL-PAIRS CORRELATION VOLUME

- This volume shows how each pixel in Image A relates to each pixel in Image B.
- It’s organized as a large 4D grid.
- Each point in this grid shows the similarity (or “correlation”) between a pixel pair from the two images.
- This is called an "all-pairs" correlation because every pixel in Image A is compared to every pixel in Image B.
- This detailed comparison helps RAFT find exact pixel matches, even in complex scenes.





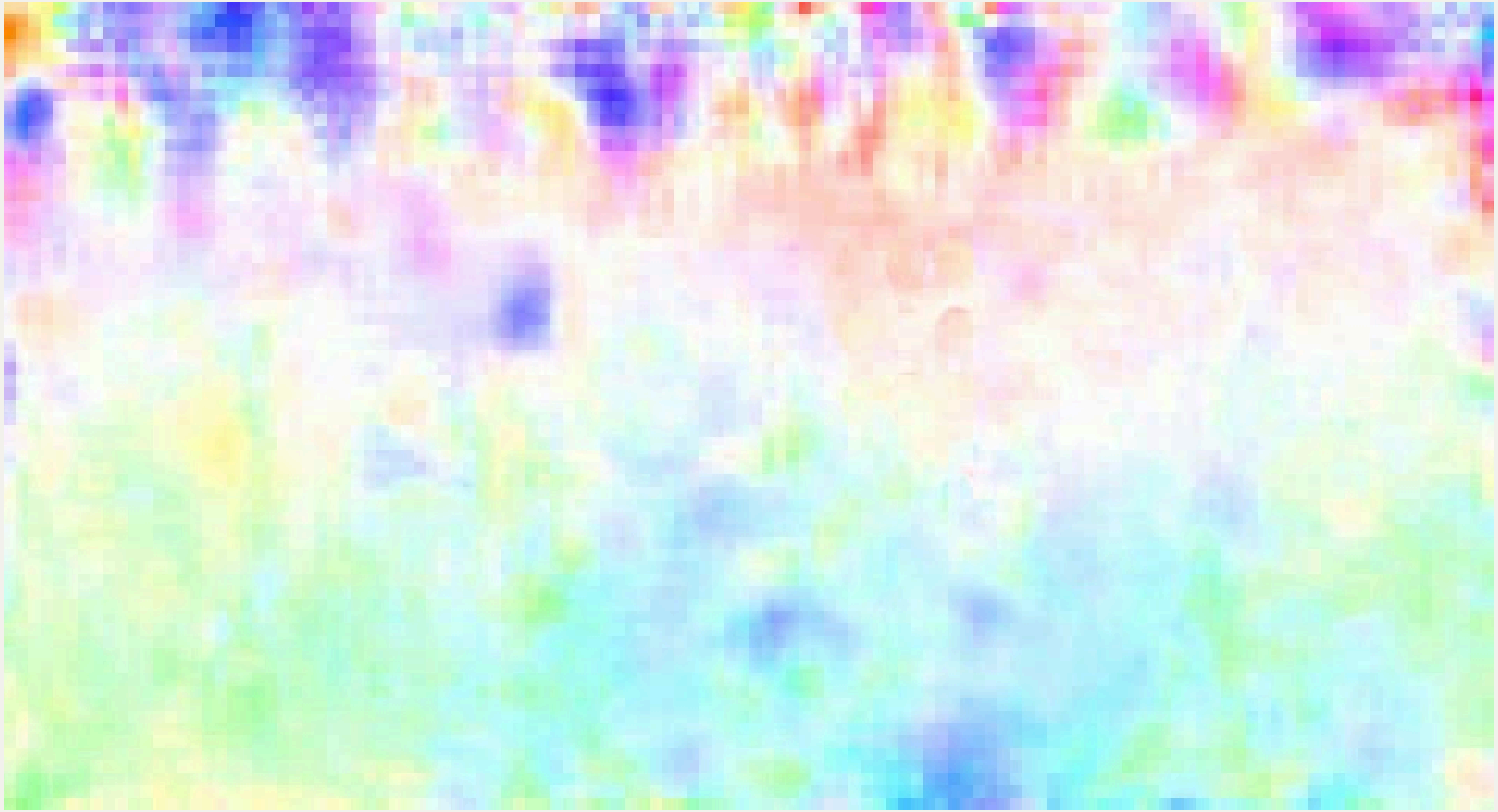
2 RAFT ARCHITECTURE



RECURRENT MOTION REFINEMENT

- RAFT uses a recurrent neural network (RNN) GRU to improve the initial flow estimates step by step.
- The GRU takes the first flow estimate and the correlation volume, then refines it through several steps.
- Each step builds on the previous one, helping the model get closer to the correct movement for each pixel.
- After several steps, the GRU produces the final flow field, which shows how each pixel in Image A moved in Image B.

DEMO



3 RESULT ANALYSIS

The RAFT model shows a sensitivity depending on the speed and clarity of motion in videos.

For slow movements, like in the previous Video, it may produce **less accurate results**, while on faster sequences with more pronounced contrasts, it provides higher-quality optical flow visualizations.

For projects requiring precision with slow movements, it may be beneficial to explore alternatives .

GENERAL CONCLUSION

5 GENERAL CONCLUSION

Conclusion

In conclusion, t**he performance of RAFT** is generally impressive, especially on data that closely matches the conditions it was trained on. It excels in handling fast-moving sequences .

However, its performance tends to decrease when dealing with slow-moving videos, where motion can be more subtle and harder to detect.

The model also requires significant **computational resources** for real-time performance.
For scenarios involving slower movements or complex scenes, adjustments like tuning parameters or using specialized models may be needed.
Despite these limitations, RAFT remains a powerful tool for motion estimation in dynamic environments.

THANK YOU FOR YOUR ATTENTION